

Marcos Vinícius de Paula Pinto

**METODOLOGIA DE ESTIMAÇÃO DE MOVIMENTO GLOBAL DE
SEQÜÊNCIAS DE VÍDEO UTILIZANDO MOVIMENTOS LOCAIS DE
PONTOS CARACTERÍSTICOS ENTRE QUADROS**

Dissertação apresentada como requisito parcial à
obtenção do grau de Mestre, no Programa de
Pós-Graduação em Engenharia Elétrica, Setor de
Tecnologia, Universidade Federal do Paraná –
UFPR.

Orientador: Marcus Vinicius Lamar

Curitiba
2005

Agradecimentos

Ao professor Marcus Vinicius Lamar, Ph.D. por seus conselhos, indicações, apontamentos e sua orientação que contribuíram para a execução deste trabalho.

À minha esposa Josi, por sua ajuda nos momentos de dificuldade e por sempre acreditar nos meus objetivos.

À minha família, meu pai Julio, minha mãe Natalina, meu irmão Fernando e minha irmã Isabela, pelo apoio e incentivo a alcançar meus objetivos.

A todos os amigos e amigas que sempre prestaram o auxílio necessário. Em especial ao Cristhian, ao Nadal, à Michelle, ao Alberto e a todos os outros companheiros do mestrado.

Aos professores do Departamento de Engenharia Elétrica da Universidade Federal do Paraná.

À Capes pela bolsa de estudos.

Sumário

Agradecimentos	ii
Sumário.....	iii
Lista de figuras.....	iv
Lista de acrônimos.....	v
Resumo	vi
Abstract.....	vii
1. Introdução	1
2. Métodos aplicados à estimação de movimento global.....	4
3. Metodologia de estimação do movimento global	10
3.1. Extração de pontos característicos	16
3.2. Detecção dos movimentos locais	23
3.3. Validação dos vetores de movimento.....	24
3.4. Cálculo dos parâmetros de transformação	28
4. Resultados.....	30
4.1. Resultado da obtenção da resposta aos cantos de Harris.....	30
4.2. Resultados da obtenção dos vetores de movimento local válidos	35
4.3. Resultados da compensação de movimento global	37
4.4. Resultado da obtenção dos parâmetros de movimento global.....	41
4.5. Análise dos resultados	55
4.6. Comparação com outros métodos.....	57
4.7. Custo Computacional.....	61
5. Conclusão.....	64
5.1. Trabalhos futuros	65
6. Referências Bibliográficas.....	66

Lista de figuras

Figura 1. Projeção da imagem na câmera com movimentação.....	10
Figura 2. Transformação da imagem com os parâmetros de zoom (1),.....	11
Figura 3. Algoritmo de casamento dos blocos (BMA).	12
Figura 4. Ponto característico e vetor de movimento local.....	14
Figura 5. Metodologia proposta.....	15
Figura 6. Idéia básica do detector de cantos de Harris.	16
Figura 7. Classificação em função dos autovalores λ_1 e λ_2	18
Figura 8. Classificação em função de R	19
Figura 9. (a) Imagem, (b) Resposta aos cantos , (c) $R >$ limiar, (d) cantos.	21
Figura 10. Invariância a escala da imagem.	22
Figura 11. Deslocamento local para as seqüências <i>stefan</i> (a), <i>bus</i> (b), <i>silent</i> (c) e <i>hall</i> (d).	24
Figura 12. Regiões de análise.....	25
Figura 13. Histograma de deslocamento horizontal de um quadrante da seqüência <i>bus</i>	26
Figura 14. Objetos de seqüências de vídeo presente no centro da imagem.....	27
Figura 15. Vetores de movimento válidos.	28
Figura 16. Resposta aos cantos de Harris da seqüência <i>bus</i>	31
Figura 17. Pontos característicos encontrados para a seqüência <i>bus</i>	32
Figura 18. Resposta aos cantos de Harris da seqüência <i>stefan</i>	33
Figura 19. Pontos característicos encontrados para a seqüência <i>stefan</i>	34
Figura 20. Vetores de deslocamento local encontrados para a seqüência <i>bus</i>	36
Figura 21. Vetores de deslocamento local encontrados para a seqüência <i>stefan</i> .36	
Figura 22. Diferença entre quadros da seqüência <i>bus</i> sem compensação de movimento global.....	38
Figura 23. Diferença entre quadros da seqüência <i>bus</i> com compensação de movimento global.....	38
Figura 24. Diferença entre quadros da seqüência <i>stefan</i> sem compensação de movimento global.....	40
Figura 25. Diferença entre quadros da seqüência <i>stefan</i> com compensação de movimento global.....	40
Figura 26. Vídeos gerados com transformações determinadas.	42
Figura 27. Translação horizontal de 1 <i>pixel</i> para a direita.	43
Figura 28. Translação horizontal de 1 <i>pixel</i> para a esquerda.....	44
Figura 29. Translação horizontal de 1 <i>pixel</i> para cima.	44
Figura 30. Translação horizontal de 3 <i>pixel</i> para a direita.	45
Figura 31. Rotação anti-horária.	47
Figura 32. Zoom In.	48
Figura 33. Zoom Out.	49
Figura 34. Zoom com translação.....	50
Figura 35. Evolução dos parâmetros de transformação da seqüência <i>stefan</i>	52
Figura 36. Evolução dos parâmetros de transformação da seqüência <i>bus</i>	53
Figura 37. Evolução dos parâmetros de transformação da seqüência <i>hall</i>	54

Figura 38.	Evolução dos parâmetros de transformação da sequência <i>silent</i>	55
Figura 39.	Translação em x da sequência <i>stefan</i> por Sikora.....	58
Figura 40.	Translação em x da sequência <i>stefan</i> detectado.	59
Figura 41.	Translação em x da sequência <i>bus</i> por Sikora.....	59
Figura 42.	Translação em x da sequência <i>bus</i> detectado.	60
Figura 43.	Escala em x e y da sequência <i>bus</i> por Sikora.....	60
Figura 44.	Escala em x e y da sequência <i>bus</i> detectados.....	61
Figura 45.	Janela de comparação do BMA	62

Lista de tabelas

Tabela 1.	Número de comparações do BMA.....	62
------------------	-----------------------------------	----

Lista de acrônimos

Sigla	Definição	Significado
MPEG	Moving Picture Experts Group	Grupo do ISO/IEC encarregado de desenvolvimento de padrões de codificação e representação de áudio e vídeo
BMA	Block Matching Algorithm	Critério de casamento de blocos através da similaridade
VO	Visual Object	Objeto visual
SVD	Single Value Decomposition	Decomposição em valor singular
DCT	Discrete Cossine Transform	Transformada discreta de cosseno
MC-DCT	Motion Compensated DCT	Transformada discreta de cosseno com compensação de movimento
MAE	Mean Absolute Error	Erro Absoluto Médio
MSE	Mean Square Error	Erro Quadrático Médio
CIF	Common Intermediate Format	Formato Intermediário Comum
3D	Three dimensional space	Espaço Tri-dimensional

Resumo

Este trabalho propõe um novo método para estimação de movimento global. Propõe-se uma junção da metodologia de extração de pontos característicos de imagens e uma metodologia de estimação de movimento global entre dois quadros consecutivos de uma seqüência de vídeo. Os pontos característicos são usados para encontrar os vetores de movimento local entre quadros. Estes movimentos locais são aplicados no método de estimação de movimento global e os parâmetros de transformação são encontrados. A vantagem desta metodologia é o ganho computacional na etapa de estimação de movimento local. Comparações deste novo método proposto são feitas com a metodologia que utiliza o algoritmo de casamento de blocos e com metodologias que utilizam técnicas baseadas no gradiente descendente a fim de validar o proposto neste trabalho.

Palavras-chaves: Estimação de movimento global e local, parâmetros de movimento.

Abstract

This work proposes a new method for global motion estimation. It is a junction of the methodology of the extraction of featuring points from images and a methodology of global motion estimation between two consecutives frames of a video sequence. The featuring points are used to find the local motion between frames. These local motions are applied in global motion estimation method and the transformations parameters are found. The advantage of this new method is to gain in computational cost in the local motion estimation part. Comparisons of the new method proposed, the BMA local motion estimation method and techniques of global motion estimation based on gradient descent method are made to validate the proposed one.

Keywords : Global and local motion estimation, movement parameters.

1. Introdução

O vídeo, do latim “eu vejo”, é uma tecnologia de processamento eletrônico de sinal que representa imagens em movimento. O termo vídeo comumente refere-se a tipos de formatos de transmissão, que pode ser digital ou analógico. A fim de diminuir a quantidade de dados transmitidos, técnicas de compressão e codificação são desenvolvidas para o vídeo digital em substituição aos vídeos transmitidos analogicamente.

Como parte importante das técnicas de compressão e codificação, a detecção de movimento tem como objetivo minimizar a quantidade de dados transmitidos diminuindo a redundância temporal que existe em uma seqüência de vídeo. Tipicamente uma seqüência de vídeo apresenta muita redundância temporal, ou seja, os quadros apresentam muita semelhança entre si, principalmente quadros subseqüentes. Para diminuir esta redundância e evitar a transmissão repetida de dados, é comum transmitir a diferença entre quadros e não os originais. O quadro anterior é utilizado como uma estimativa do quadro atual, e os resíduos, a diferença entre o estimado e o valor real, são codificados. Quando a estimativa é boa, é mais eficiente codificar o quadro residual ao original.

Considerando o fato de tipicamente o vídeo ser uma visão de objetos em movimento, é possível melhorar a predição estimando o movimento de cada bloco da imagem entre o quadro anterior e o atual como aplicado nos padrões MPEG-1 [1] e MPEG-2 [2]. Os quadros das seqüências de vídeo destes padrões de codificação são divididos em macro-blocos de 16x16 *pixels*. Cada macro-bloco tem seu par correspondente com maior semelhança no quadro subseqüente e somente o vetor de movimento deste macro-bloco e os resíduos (diferença) são codificados, diminuindo muito a quantidade de dados na transmissão do vídeo e aumentando a eficiência.

Para melhorar ainda mais a codificação, as novas técnicas de compressão de vídeo, como o MPEG-4 [3], propõem a divisão da seqüência de vídeo em objetos visuais (VO). Os VO's são entendidos como partes que compõem uma imagem. Como exemplo, podemos citar o plano de fundo

(background), o objeto em foco, áudio de fundo (background áudio), o áudio principal, entre outros. Com a separação da sequência em VO's, é possível aplicar codificações diferentes para cada um melhorando a eficiência na codificação.

Para a obtenção do objeto "plano de fundo", por exemplo, é necessário conhecer o movimento que a imagem de fundo faz em relação aos demais objetos da cena. Esta movimentação se dá devido à movimentação e operação da câmera de aquisição, sendo conhecido como movimento global da imagem. Algumas técnicas apresentadas por outros autores, para a estimação da movimentação global da sequência de vídeo, utilizam gradiente descendente com minimização de uma função custo [4], outros também utilizam a teoria de fluxo óptico [5] ou também, informações espaço-temporal do vídeo [6]. Ainda existem proposições de utilização de redes neurais [7] ou mesmo métodos estatísticos [8] na estimação do movimento global.

As técnicas de estimação que utilizam gradiente descendente para minimização da função custo têm um grande peso computacional por possuírem um processo iterativo bastante custoso. Outras técnicas que fazem uso da informação espaço-temporal para estimação do movimento global já são bem menos custosas do ponto de vista computacional.

Este trabalho apresenta uma adaptação da técnica de estimação de movimento global por informação espaço-temporal. É composta basicamente de duas etapas. A primeira etapa se constitui na detecção dos movimentos locais da imagem e a segunda na estimação do movimento global. É uma técnica de fácil implementação e com bons resultados. O proposto aqui é substituir a etapa de busca de movimento local utilizada por [6] pela estimação de movimento de pontos característicos da imagem contida no quadro e posteriormente aplicar na etapa de estimação dos parâmetros de movimento.

Neste trabalho, são apresentados no primeiro capítulo a metodologia proposta de estimação do movimento global de uma sequência de vídeo, com demonstração das técnicas de extração de pontos característicos, detecção de movimento local e cálculo dos parâmetros do movimento.

No capítulo seguinte é apresentada uma análise da técnica com avaliação dos resultados obtidos em seqüências de vídeo padrão e seqüências de vídeo sintetizadas com movimentos conhecidos.

Por fim, é apresentada uma conclusão do trabalho e também possíveis aplicações extras desta técnica, juntamente com proposições de trabalhos futuros.

2. Métodos aplicados à estimação de movimento global

Neste capítulo serão apresentados alguns trabalhos na área de estimação de movimento global de seqüências de vídeo que encontramos na literatura e que foram utilizados como embasamento nos estudos desta área da codificação de vídeo. O objetivo principal deste capítulo é mostrar o estado da arte na detecção de movimento global a fim de se conhecer os trabalhos e compará-los com o proposto. Vários métodos de detecção dos parâmetros de transformação são propostos e utilizados. Alguns destes métodos são baseados em gradiente descendente para a obtenção dos parâmetros de transformação que representem o movimento global do vídeo, outros ainda utilizam a teoria de fluxo óptico, ou mesmo, fazem o uso dos vetores detectados para compensação de movimento de algoritmos de codificação, dentre outros que podemos encontrar.

Dufaux e Konrad [4] propõem um método de estimação global de movimento, que utiliza um algoritmo baseado em uma pirâmide de imagens e gradiente descendente. A pirâmide é composta de dois quadros subseqüentes, com três resoluções diferentes cada quadro. O gradiente descendente é montado a partir das derivadas parciais dos erros da estimação do quadro em relação ao quadro real, pela variação de cada parâmetro de transformação. O processo inicia com a estimação dos valores da translação, por BMA (algoritmo por procura de bloco) em três passos, com os quadros de menor resolução. Após a estimação da translação, é utilizado um algoritmo de gradiente descendente, com os quadros de menor resolução, para a estimação e o refinamento de todos os parâmetros de movimento global. Os parâmetros de movimento global que foram encontrados são re-escalados, para a nova resolução dos quadros da etapa seguinte. Nas outras duas etapas o processo é o mesmo, exceto pelo BMA, com gradiente descendente seguido de re-escalamento dos parâmetros de movimento global. Outra proposta importante deste autor é referente à estimação do erro, no gradiente descendente. Os erros, *pixel a pixel*, acima de um limiar não são

considerados, pois podem fazer parte do movimento de objetos e não do fundo da imagem.

Kim e Choi [6] propuseram um método de estimação de movimento global que utiliza modelo de seis parâmetros e informação espaço-temporal do vídeo. Primeiramente calculam vetores de movimento local para blocos de 16 x 16 pixels usando o algoritmo BMA. Depois estimam os seis parâmetros usando método dos mínimos quadrados. Os vetores que não fazem parte do movimento global são retirados e inicia-se um processo iterativo até os parâmetros convergirem. Depois da estimativa, uma decisão de utilização de compensação de movimento é feita porque nem sempre o movimento da câmera existe no vídeo. Para esta decisão é utilizado o coeficiente de múltipla determinação, que é usualmente utilizado para teste de adequação de modelo em análises estatísticas. A estimativa do movimento global é utilizada para a detecção e segmentação dos objetos contidos no vídeo.

Smolic, Sikora e Ohm [9] apresentam uma forma diferente de estimação global de movimento que possui três etapas. Cada etapa tem como função refinar os parâmetros de movimento e aumentar gradativamente a complexidade do modelo de movimento global. Na primeira etapa é estimada a translação entre o quadro atual e o quadro estimado do mosaico, através de BMA. Após a estimação da translação vem a segunda etapa, onde são estimados os parâmetros de zoom, rotação e os parâmetros de translação são refinados. O processo da etapa dois é iterativo e baseado no fluxo ótico, campo de vetores de movimento. O primeiro passo desta etapa é a reconstrução do quadro atual, a partir do mosaico e dos parâmetros de movimento estimados, somente translação, na primeira etapa. Com o quadro reconstruído são estimados novos vetores de movimento em relação ao quadro atual. Com os novos vetores de movimento são estimados os parâmetros de movimento global, através de um sistema de equações. Este processo da etapa dois fica em *loop*, enquanto a diferença entre o quadro reconstruído e o quadro atual for maior que um limiar estabelecido. A etapa três é semelhante à etapa dois, porém com o modelo de transformação parabólico, que leva em conta transformações de mais alta ordem.

Wren [10] propõe um método de estimação dos parâmetros de movimento global com o auxílio do usuário, que encontrará pontos similares em diferentes figuras. Este processo não é aplicado especificamente para vídeo. O algoritmo de estimação dos parâmetros movimento é baseado na solução de um sistema de equações, composto do deslocamento de cada ponto, escolhido pelo usuário em função dos parâmetros de movimento global. Para a resolução do sistema de equações é utilizado o método SVD (decomposição em valor singular). O artigo não propõe nenhum método de estimação de movimento por procura de bloco, para formar o sistema de equações. Porém tem como vantagem a solução do sistema de equações de uma forma mais simples e computacionalmente menos penosa que o gradiente descendente, por exemplo.

Outro trabalho importante, na parte de estimação dos parâmetros de movimento global, é o de Zhu, Xue e Ong [11]. Neste trabalho, é utilizado um método de extração dos parâmetros de movimento global, através dos vetores de movimento. Neste caso, os vetores são obtidos de vídeos codificados pela transformada discreta dos co-senos, com compensação de movimento (MC-DCT). Com a ajuda dos coeficientes da DCT (transformada discreta de co-senos), são extraídos os vetores de movimento espúrios, sem uma reconstrução explícita da imagem. Os vetores são usados no cálculo dos parâmetros de movimento global, através de mínimos quadrados.

Mais um trabalho em estimação de movimento global é apresentado por Kim e Kim [12]. Eles propõem um método de estimação dos parâmetros, onde é realizada inicialmente uma seleção dos blocos. Os blocos selecionados são os que apresentam bordas ou altas texturas. Com os blocos selecionados, é feita a procura entre dois quadros consecutivos. Através dos deslocamentos estimados pela procura de blocos, são calculados os parâmetros de movimento, por mínimos quadrados. Os parâmetros de movimento global são utilizados na etapa de filtragem. Na filtragem são retirados os valores de deslocamento que estão excessivamente fora da tendência central estimada. Após a etapa de filtragem, é feita uma nova estimação dos parâmetros de movimento global. O método fica

neste *loop* de cálculo dos parâmetros de movimento global e filtragem enquanto houver alguma modificação significativa.

Giunta e Mascia [5] mostram outra forma de estimação de movimento global, baseada em fluxo ótico. O primeiro passo do método é a estimação dos vetores de movimento, através de BMA (algoritmo de procura por bloco), com função de avaliação MAE (erro médio absoluto). O propósito do primeiro passo é minimizar os MAE's entre os blocos do quadro de referência e o quadro posterior. Encontrados os deslocamentos horizontais e verticais são montadas as equações de movimento global, com modelo parabólico, para cada bloco. Para avaliar a qualidade de cada equação de movimento são calculados os pesos, de cada expressão. Os pesos levam em conta a variação do erro médio absoluto pela variação do deslocamento dos blocos. Com os pesos de cada expressão calculados e as expressões de movimento dos blocos, são estimados os parâmetros de movimento global, através da solução do sistema de equações.

Wang e Wang [13] apresentam uma forma de estimação de movimento global semelhante à de outros trabalhos. Inicialmente é feita uma procura por similaridade de bloco, para estimar o campo de vetores de movimento. Do campo de vetores de movimento são calculadas as derivadas parciais espaciais dos vetores de movimento. Com as derivadas parciais espaciais de cada vetor calculadas são feitas as filtrações. As filtrações equivalem à retirada de algumas expressões, da avaliação do movimento global. O critério de retirada dos vetores de movimento é referente à média das derivadas parciais espaciais dos vetores de movimento. Caso o módulo da diferença entre a derivada parcial espacial do vetor de movimento e o valor médio das derivadas parciais espaciais for maior do que o limiar o vetor não é considerado na estimação de movimento. Uma particularidade deste método é a utilização do valor médio das derivadas parciais espaciais dos vetores de movimento, após a filtração, para o cálculo dos parâmetros de movimento.

Outra forma de estimar os parâmetros de movimento é apresentada por Chen, Lin e Chen [7], onde são utilizadas redes neurais para estimar o movimento global. O processo proposto por eles possui duas etapas. A primeira

etapa consiste em uma procura por *pixels* semelhantes, entre dois quadros consecutivos, utilizando uma rede neural do tipo Hopfield. As redes neurais Hopfield são redes realimentadas, com treinamento não supervisionado e tem como objetivo minimizar uma função de energia. A função de energia tenta retratar numericamente a similaridade entre os blocos de *pixels* testados. Na segunda etapa, com os deslocamentos calculados na primeira etapa, são estimados os parâmetros de movimento. O cálculo dos parâmetros é feito através de uma rede neural *feed-forward* de três camadas, com treinamento supervisionado. O treinamento é realizado com seqüências de vídeo artificiais, em que os parâmetros de movimento já são conhecidos. Este artigo não apresenta especificamente um método para estimação de movimento global, movimento do fundo da seqüência de vídeo, mas sim um método de estimação de movimento rígido qualquer em três dimensões.

Cheung e Siu [14] apresentam um estimador de movimento global para a construção de imagens panorâmicas de fundo de uma seqüência de vídeo. O estimador de movimento proposto neste artigo tem como inovação a utilização de diversas referências para a estimação de movimento global. Em métodos anteriores é sempre utilizado o mosaico ou os parâmetros são concatenados quadro a quadro. Neste caso, o quadro de referência é modificado em duas situações. A primeira situação é quando a diferença, quadros visualmente diferentes, entre o quadro de teste e quadro de referência forem muito grandes. A outra situação de mudança do quadro de referência é quando a diferença, em número de quadros, entre o quadro de teste e o quadro de referência for muito grande. Para o cálculo dos parâmetros de movimento global é utilizada procura por bloco seguida de gradiente descendente.

Analisando as metodologias existentes para detecção do movimento global em uma seqüência de vídeo, podemos notar que em sua grande maioria, o casamento de blocos para detecção de vetores de movimento local predomina. Também o algoritmo que utiliza o gradiente descendente é bastante utilizado pelos autores. Novas metodologias que utilizam redes neurais e lógica *fuzzy*, tentam

melhorar tanto a velocidade quanto a qualidade da estimação de movimento global.

O artigo de Dufaux é base de estudos para muitos autores. Os resultados obtidos por ele são bastante precisos, porém, faz uso da metodologia que utiliza gradiente descendente, o que torna o processo de detecção bastante custoso do ponto de vista computacional.

Para melhorar a velocidade da estimação, outros autores estudados propõem a utilização dos próprios vetores de movimento que são detectados para compensação de movimento das codificações de vídeo. Porém, muitas vezes a detecção destes vetores não é feita a fim de representar a movimentação da imagem e sim para melhorar a estimativa futura do bloco e diminuir a quantidade de dados transmitida.

Este trabalho é baseado nas metodologias de Kim e Choi e também na metodologia de Wren. Consiste em substituir a etapa de estimação de movimento local de Kim pela utilizada por Wren. Observando que o trabalho de Wren é feito para fotos e não para vídeo, é necessário alterar a etapa de casamento de pontos por uma mais automatizada, sem necessidade da intervenção do usuário.

No próximo capítulo será demonstrado o método de estimação de movimento global proposto neste trabalho. Serão abordadas as técnicas de estimação de movimento local, cálculo dos parâmetros de transformação e resultados obtidos com análises em seqüências de vídeos.

3. Metodologia de estimação do movimento

global

A movimentação da câmera de vídeo ($C \rightarrow C'$) gera uma transformação da imagem (M) de um quadro em relação ao quadro anterior devido à projeção (m, m') deste objeto no plano de detecção da câmera (R).

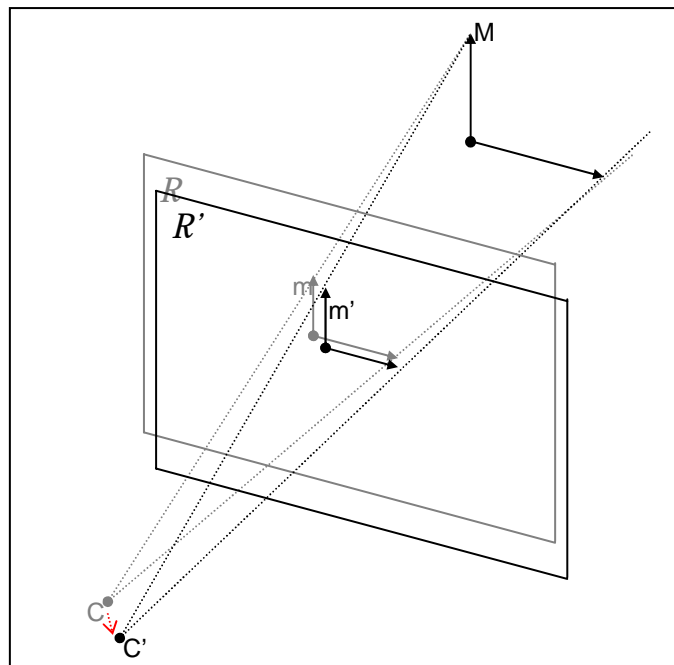


Figura 1. Projeção da imagem na câmera com movimentação

Para descrever esta transformação existem vários modelos [16]. Alguns desses modelos são: o modelo de projeção geométrica de imagens, o modelo de perspectiva de movimento finito de 6 e 8 parâmetros, o modelo de pseudoperspectiva de movimento, e outros. Neste trabalho o modelo escolhido para implementação é o de perspectiva de movimento finito com 8 parâmetros por representar quase todos os movimentos possíveis de uma câmera de vídeo com alta fidelidade e por apresentar grande simplicidade matemática [6].

Este modelo é definido como

$$X = \frac{ax + by + c}{gx + hy + 1} \quad (1)$$

$$Y = \frac{dx + ey + f}{gx + hy + 1} \quad (2)$$

onde a, b, c, d, e, f, g e h são os parâmetros de movimento da câmera, x e y são as coordenadas de um ponto da imagem e X e Y as coordenadas deste mesmo após a transformação.

Nas figuras a seguir é mostrada a transformação da imagem para zoom (a e e), translação (parâmetros c e f), rotação (b, d, g e h) e inclinação (b e d) a fim de ilustrar os movimentos da câmera contemplado por este modelo.

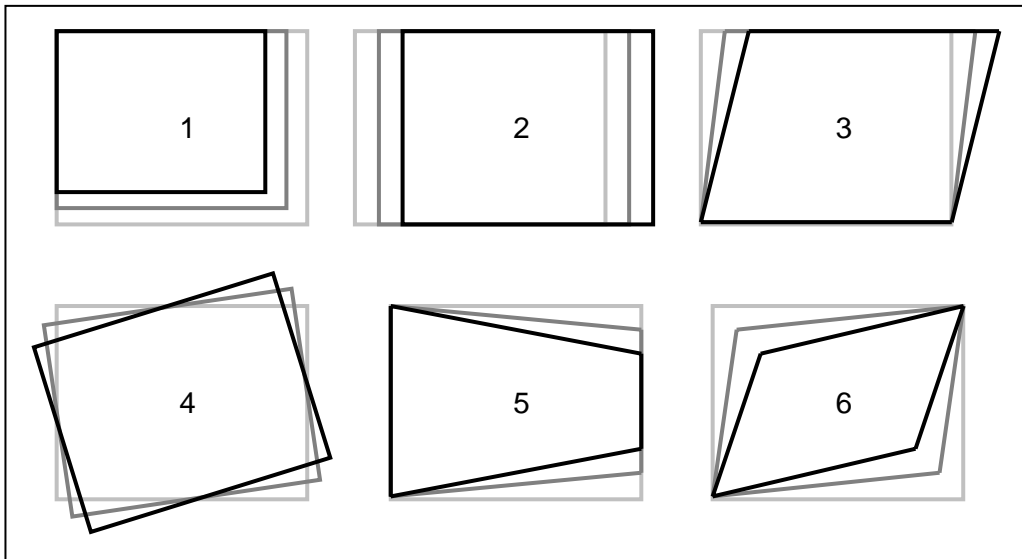


Figura 2. Transformação da imagem com os parâmetros de zoom (1), translação (2), inclinação (3), rotação (4), *pan* (5) e livre (6).

Para a estimação destes parâmetros de movimento é necessário detectar alguns vetores de movimento local ocorrido por certos pontos na imagem para aplicarmos nas equações e obtermos os valores dos parâmetros.

Primeiramente foi estudada a possibilidade de utilização do método de obtenção dos vetores de movimento local utilizado na compensação de movimento dos padrões de compressão de vídeo MPEG-1 e MPEG-2, como é feito por Kim. Consiste em segmentar a imagem em blocos de tamanho definido, tipicamente 16×16 *pixels*, e em seguida cada bloco encontrado no quadro é comparado com todos os pontos de uma janela de busca no quadro anterior, localizado nas mesmas coordenadas do bloco quadro original.

Este método é mostrado na Figura 3.

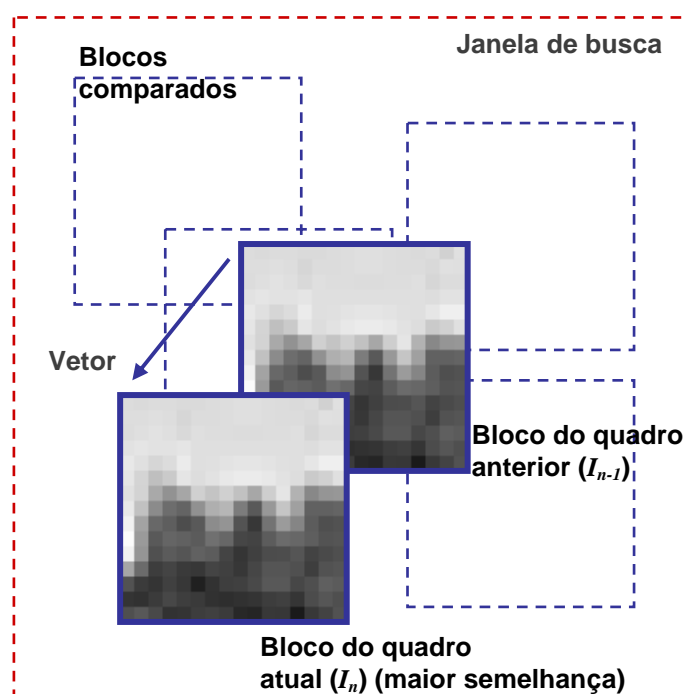


Figura 3. Algoritmo de casamento dos blocos (BMA).

As coordenadas do ponto onde houver o melhor casamento dos blocos menos as coordenadas do ponto de origem do bloco no quadro original é considerado o vetor de movimento local do bloco. Como este método busca a melhor semelhança para diminuir a quantidade de dados transmitidos, é possível que o vetor encontrado não represente o movimento real do bloco e sim a posição de maior semelhança, o que posteriormente atrapalharia a estimação do movimento global. Uma observação que deve ser feita é que a quantidade de

cálculo torna o método muito custoso computacionalmente, pois a quantidade de blocos analisada é alta.

Como exemplo, supondo uma janela de busca e o bloco com tamanhos de 16×16 *pixels*, pelo algoritmo BMA serão feitas 65536 comparações entre *pixels* para cada bloco da imagem. Como numa imagem de 352×288 *pixels* (CIF) são 22×18 blocos, total de 396 blocos, a quantidade de comparações pode chegar a aproximadamente 25 milhões para o caso de busca completa (*full search*) do bloco. Se utilizarmos um método de busca rápida [17] como a busca em 3 passos (*3 step search*), ou a por diamante (*diamond search*), a quantidade de comparações pode diminuir consideravelmente, mas o resultado da busca do movimento passa a não ser necessariamente o movimento real do bloco.

Para melhorar esta etapa, é proposta a substituição do BMA pela detecção de movimento de pontos característicos em cada quadro da seqüência de vídeo. É uma técnica conhecida e utilizada na construção de imagens panorâmicas a partir de fotos do ambiente. Consiste basicamente em detectar pontos importantes como cantos de objetos, cruzamentos de retas, em dois quadros subseqüentes e detectar a diferença de posição em que se encontram estes pontos em cada quadro. O resultado é o vetor de movimento do ponto (movimento local). Um exemplo é mostrado na Figura 4, onde o mesmo ponto é encontrado em dois quadros subseqüentes de uma seqüência de vídeo (I_{n-1} e I_n). O deslocamento deste ponto característico entre um quadro e outro representa o vetor de movimento local.

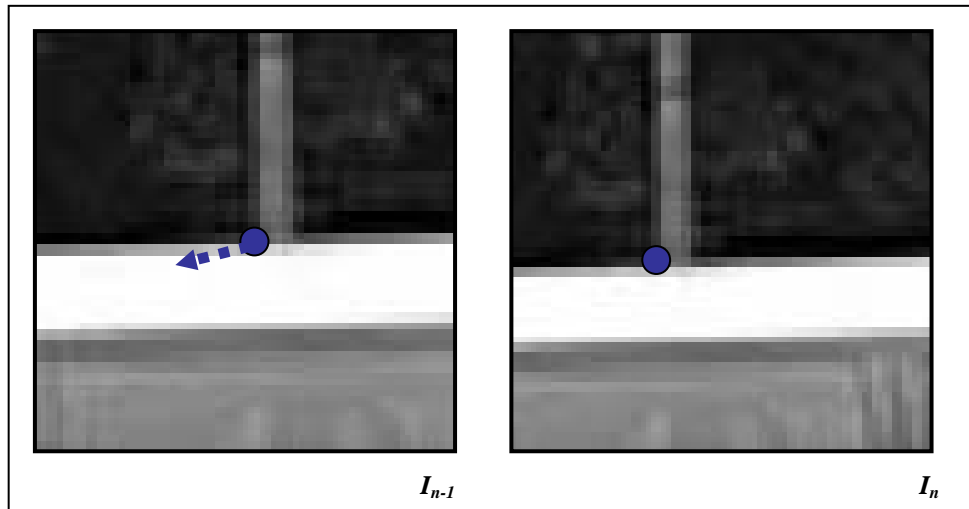


Figura 4. Ponto característico e vetor de movimento local.

Como a quantidade de pontos encontrados é bem menor que a quantidade de blocos da metodologia anterior, e como a comparação entre os pontos é feita apenas com pontos próximos, a quantidade de comparações é muito menor, resultando em um custo computacional gasto nesta etapa bem baixo.

Com todos os movimentos locais detectados, são então selecionados os vetores válidos, aqueles que realmente representam a movimentação da câmera, para serem utilizados na etapa de estimação dos parâmetros de transformação (movimento global).

A Figura 5 mostra um resumo da metodologia proposta.

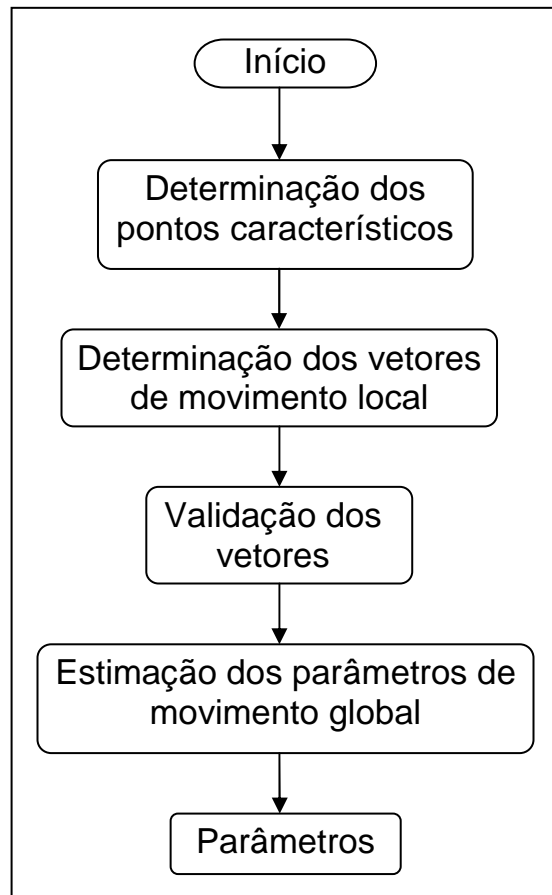


Figura 5. Metodologia proposta.

Nas seções seguintes são apresentados detalhes da metodologia proposta e no capítulo seguinte os resultados obtidos em testes e simulações.

3.1. Extração de pontos característicos

Esta é a primeira etapa de obtenção dos parâmetros de movimento global de uma seqüência de vídeo. Consiste basicamente em extrair os pontos característicos (cantos, pontos, cruzamentos de retas, etc.) de cada quadro do vídeo. A extração de pontos característicos também é utilizada em aplicações como alinhamento de imagem, reconstrução 3D, reconhecimento de objetos, navegação robótica, entre outras.

A idéia básica do detector de cantos de Harris [18] consiste em definir um pequeno quadrado dentro da imagem e deslocar em várias direções. Estas variações de deslocamento provocam pequenas variações de intensidade neste quadro. Dependendo do tipo de variação de intensidade, este quadro contém um canto, uma borda ou mesmo uma região plana. A Figura 6 mostra esta característica.

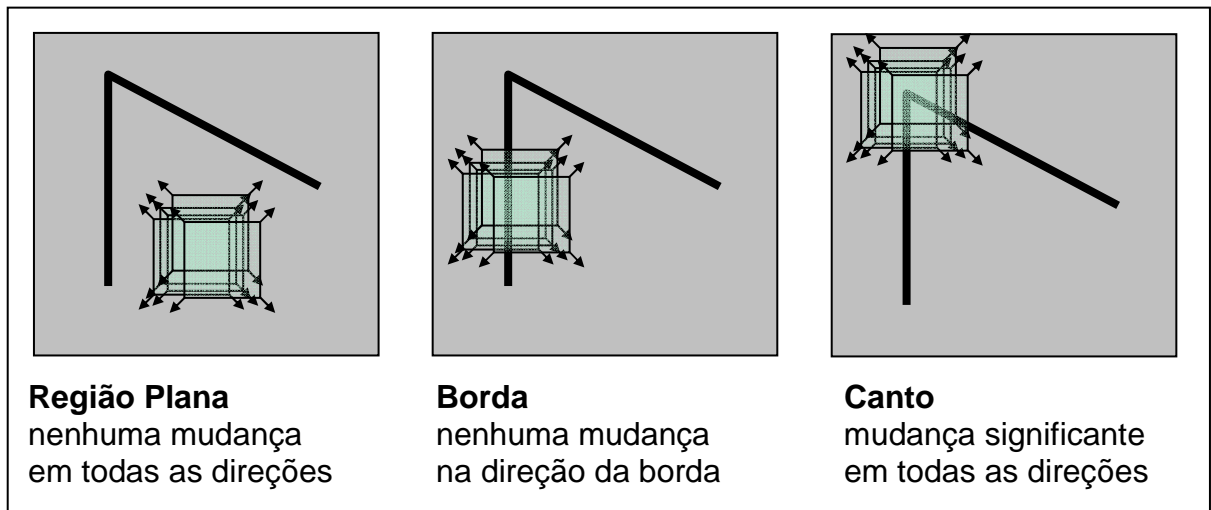


Figura 6. Idéia básica do detector de cantos de Harris.

A especificação matemática é demonstrada como

$$E(x, y) = \sum w(u, v) \cdot |I_{x+u, y+v} - I_{u, v}|^2 \quad (3)$$

onde I é a intensidade da imagem, E a variação de intensidade introduzida pelo pequeno deslocamento da imagem e w especifica o tipo de janela do quadro, sendo a unidade na região retangular especificada e zero nas demais regiões.

Se a janela considerada for pequena (2x2 *pixels*), então é possível simplificar por uma aproximação bilinear

$$E(u, v) = [u, v] \cdot M \begin{bmatrix} u \\ v \end{bmatrix} \quad (4)$$

onde M é uma matriz 2x2 computada a partir das derivadas da imagem

$$M = \sum W(x, y) \cdot \begin{bmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} & \left(\frac{\partial I}{\partial y}\right)^2 \end{bmatrix} \quad (5)$$

sendo $I(x, y)$ a intensidade da imagem considerada.

O cálculo das derivadas parciais da imagem é feito da seguinte forma

$$\frac{\partial I}{\partial x} = I(x, y) \otimes (-1, 0, 1) \quad (6)$$

$$\frac{\partial I}{\partial y} = I(x, y) \otimes (-1, 0, 1)^T \quad (7)$$

ou seja, a convolução com um filtro derivativo horizontal e vertical. Harris ainda propõe que $\left(\frac{\partial I}{\partial x}\right)^2$, $\left(\frac{\partial I}{\partial y}\right)^2$ e $\frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y}$ sejam filtrados utilizando uma janela circular gaussiana para diminuir o ruído causado pela janela binária e retangular das derivadas.

Segundo o autor, os autovalores (λ_1 e λ_2) da matriz M , representam a característica da região analisada. Quando os dois autovalores são pequenos, temos uma região "plana". Se um dos autovalores for maior que o outro, então

temos uma borda da imagem. Se os dois autovalores forem grandes e próximos temos um canto da imagem (Figura 7).

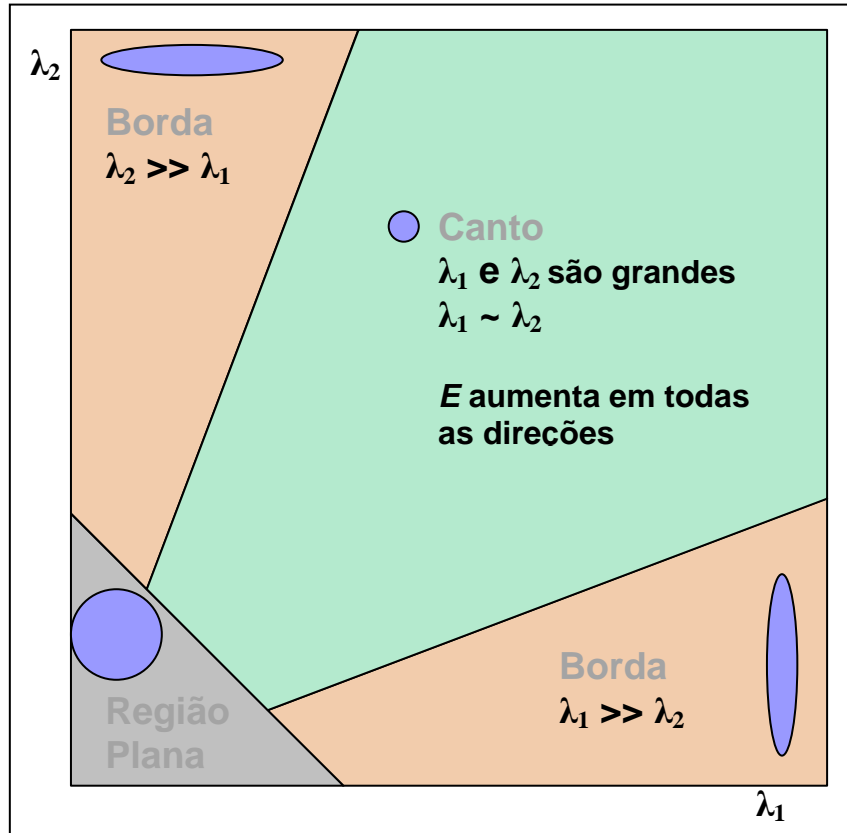


Figura 7. Classificação em função dos autovalores λ_1 e λ_2

Harris propôs a utilização de uma função chamada de "resposta aos cantos" para evitar a decomposição dos autovalores da matriz M , definida como

$$R = \det\{M\} - k \cdot (\text{trace}\{M\})^2 \quad (8)$$

$$\det\{M\} = \lambda_1 \cdot \lambda_2 = \left(\frac{\partial I}{\partial x}\right)^2 \cdot \left(\frac{\partial I}{\partial y}\right)^2 - \frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \quad (9)$$

$$\text{trace}\{M\} = \lambda_1 + \lambda_2 = \left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2 \quad (10)$$

onde k é uma constante empírica de valores típicos proposto por Harris entre 0,04 e 0,06. Então, dependendo do valor de R temos a classificação da região da imagem em região plana, canto ou borda.

Se a resposta da função for pequena, independente se positiva ou negativa, será classificado com região plana. Se a resposta for negativa grande, então será classificado como borda. E se a resposta for positiva grande será um ponto característico.

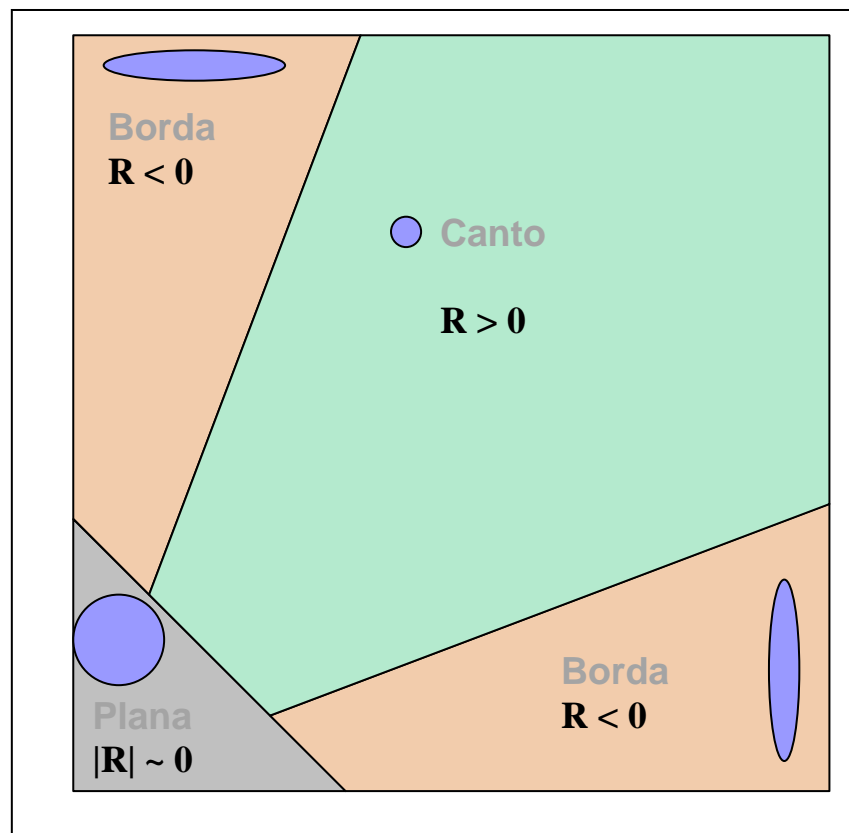


Figura 8. Classificação em função de R

Para detectarmos os pontos de canto da imagem (pontos característicos) é necessário encontrar os pontos com alto valor de resposta aos cantos. Apenas as regiões acima de um limiar são analisadas ($R > \text{limiar}$) a fim de eliminar a possibilidade de encontrar pontos característicos em regiões planas. Para a determinação da posição precisa dos pontos característicos, é proposto

determinar os máximos locais do resultado como sendo a posição dos pontos. Para a determinação dos máximos locais, uma janela de tamanho fixo é analisada e o máximo global desta janela é julgado como sendo um máximo local da imagem. Esta janela é deslocada em toda a imagem a fim de encontrar todos os máximos locais (pontos característicos).

Na Figura 9 é mostrado um exemplo da detecção de pontos característicos. A partir da imagem original (a) é obtida a resposta aos cantos desta imagem (b). Em seguida, apenas os valores de R maiores que um limiar são analisados (c), e por fim, os máximos locais encontrados são determinados como sendo pontos característicos (d).

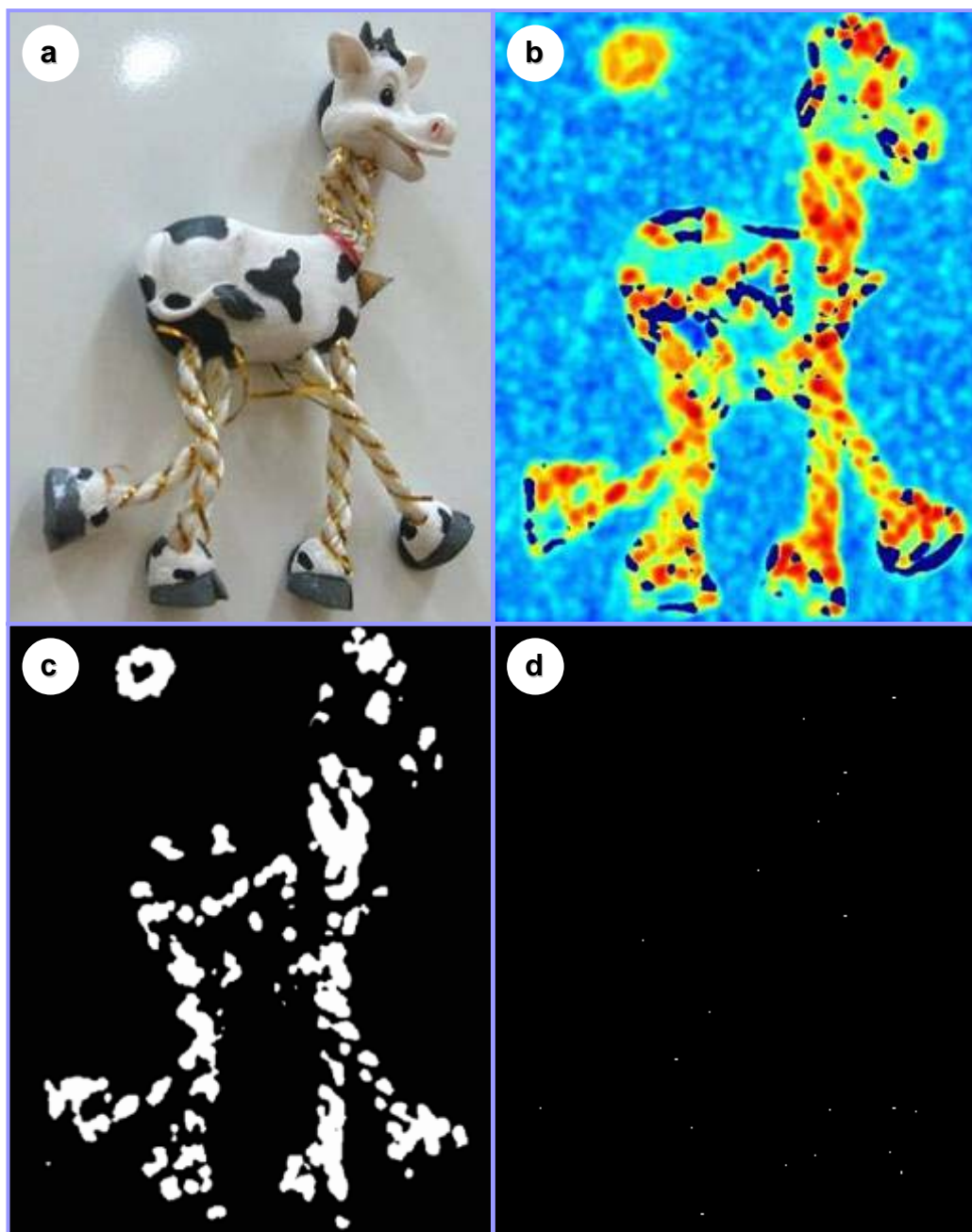


Figura 9. (a) Imagem, (b) Resposta aos cantos , (c) $R > \text{limiar}$, (d) cantos.

Este método de detecção de pontos característicos é invariante à rotação da imagem, parcialmente invariante às mudanças de intensidade e não-invariante à escala na imagem (zoom). Estas são características importantes, pois nas seqüências de vídeo são encontradas mudanças de luminosidade e também mudança de escala (zoom), o que, em princípio, inviabilizaria sua utilização.

Porém, se admitir-mos que uma seqüência de vídeo tem uma alta taxa de atualização de quadros (geralmente 30 quadros por segundo), então não ocorrerão mudanças significativas de luminosidade e escala (zoom) entre dois quadros subseqüentes, apenas mudanças sutis. Isto faz com que praticamente os mesmos pontos sejam encontrados em dois quadros seqüenciais, o que não ocorreria se a análise fosse feita entre quadros muito distantes.

A Figura 10 mostra que se houvesse uma alteração grande na escala da imagem, a região que seria detectada como canto passa a ser estimada com bordas (existe uma variância quanto à escala). Porém, quando a variação entre os quadros é pequena, a região continua sendo detectada como canto (invariância quanto à escala). O mesmo pode ser verificado para mudança de intensidade (brilho) da imagem.

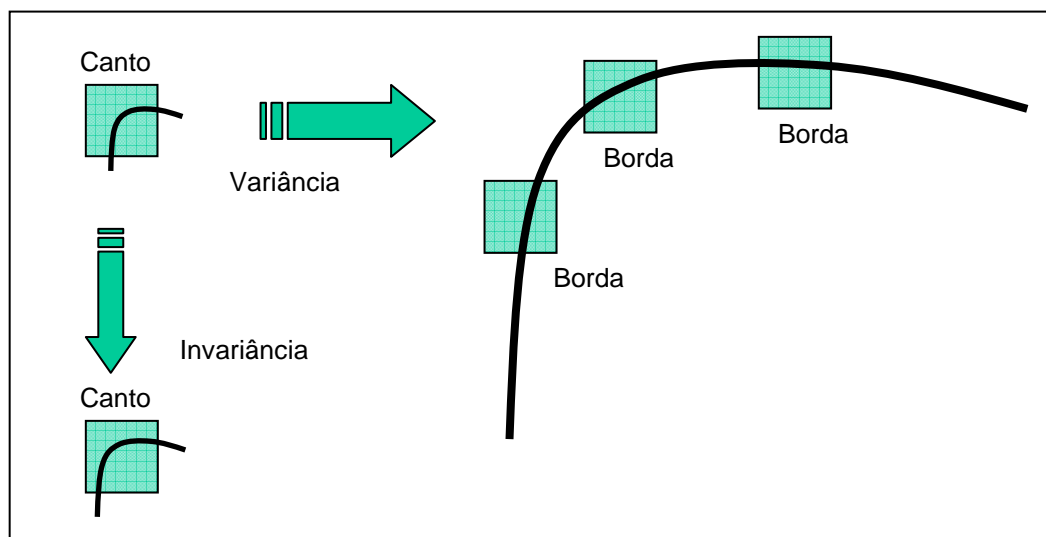


Figura 10. Invariância a escala da imagem.

Após encontrarmos os pontos principais nos quadro atual e anterior do vídeo é necessário encontrar os vetores de movimento local (movimentação dos pontos) do vídeo. Estes vetores representam a transformação da imagem e são utilizados para estimar os parâmetros de movimento.

3.2. Detecção dos movimentos locais

Após encontrarmos os pontos principais nos quadro atual e anterior do vídeo é necessário encontrarmos os vetores de movimento local que representam a transformação da imagem de acordo com os parâmetros.

Primeiramente um bloco de tamanho definido, centrado no ponto encontrado na imagem, é comparado com blocos também de tamanhos definidos, centrados em pontos encontrados no quadro seguinte, desde que respeitem um critério de distância. Este critério é devido à baixa movimentação local dos pontos, pois a taxa de aquisição do vídeo é alta. Como distância máxima permitida para a análise, foi admitida sendo de 16 *pixels*, ou seja, somente um deslocamento de até 16 *pixels* em qualquer direção é analisado.

A semelhança entre os blocos é calculada através da correlação cruzada

$$C = \sum_i \sum_j (I(x-i, y-j) - \bar{I})(I'(x'-i, y'-j) - \bar{I}') \quad (11)$$

onde I e I' são o valor de intensidade de certo ponto de um quadro e do quadro seguinte, e \bar{I} e \bar{I}' a média da intensidade da vizinhança considerada. A distância entre os dois pontos com maior correlação é considerada como sendo o vetor de movimento local ocorrido na imagem. Isto é feito para todos os pontos encontrados na imagem atual.

A Figura 11 mostra estes vetores de deslocamento local encontrados para as seqüências de vídeo *stefan*, *bus*, *silent* e *hall*.



Figura 11. Deslocamento local para as seqüências *stefan* (a), *bus* (b), *silent* (c) e *hall* (d).

Notamos a necessidade de a imagem conter algumas altas frequências, pois em caso negativo, a detecção dos vetores fica prejudicada, ocasionando um número pequeno de vetores encontrados.

3.3. Validação dos vetores de movimento

Para evitar erros no cálculo dos parâmetros de transformação, alguns vetores de movimento encontrados que não fazem parte do movimento de câmera, como os movimentos do objeto em foco, não devem ser considerados. Podemos notar pela Figura 11 que a grande maioria do movimento encontrado é

provocada pela câmera de vídeo, porém alguns movimentos detectados são representativos dos movimentos do objeto que se encontra em foco.

Por exemplo: na seqüência *stefan*, são encontrados alguns vetores de movimento do jogador de tênis; na seqüência *bus* são encontrados vetores para o movimento do ônibus; na seqüência *hall*, são encontrados movimentos para as duas pessoas apesar de o vídeo ser estático; na seqüência *silent* também é encontrado vetores de movimento para a mão da pessoa e também para a movimentação da cabeça do locutor.

Para podermos validar os vetores que farão parte do cálculo dos parâmetros, a imagem é subdividida em 5 regiões de análise (Figura 12), os quatro quadrantes da imagem mais a região central.

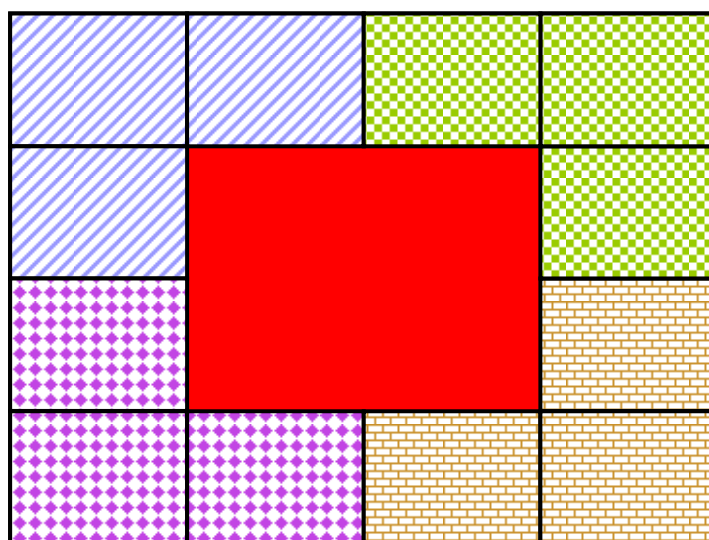


Figura 12. Regiões de análise

Dividindo a imagem em regiões, podemos determinar uma tendência de movimento para cada uma. Para efeito de validação dos vetores, somente os que acompanharem esta tendência serão considerados.

A tendência de movimento de cada região é encontrada determinando-se a moda (maior valor do histograma de movimento) dos vetores encontrados em cada região. Após a determinação desta moda, somente os vetores que tiverem uma distância mínima deste valor são considerados.

Para os casos analisados, foi verificado que uma distância de 2 *pixels* da tendência de movimento gera bons resultados.

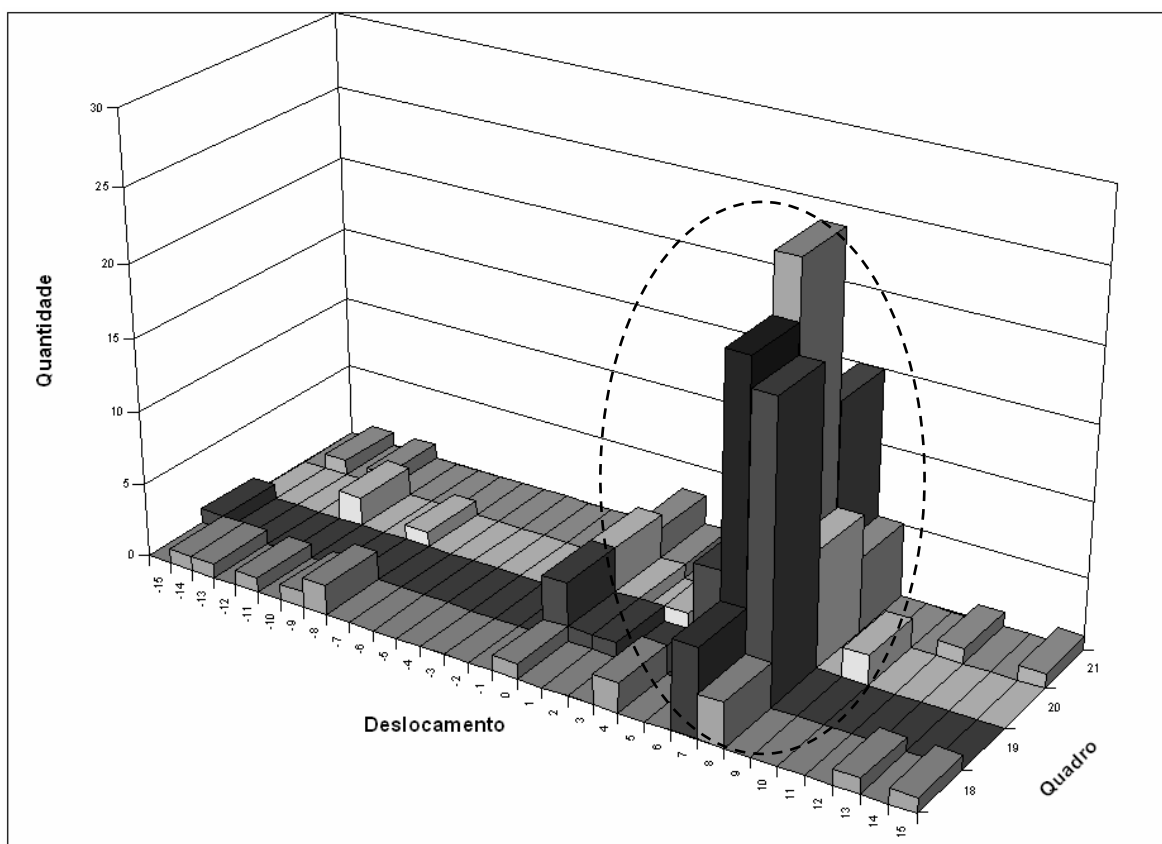


Figura 13. Histograma de deslocamento horizontal de um quadrante da seqüência *bus*.

A Figura 13 mostra o histograma de deslocamento horizontal do quadrante superior esquerdo dos quadros 18, 19, 20 e 21 da seqüência *bus*. Fica evidenciada a maior ocorrência de deslocamento horizontal em 7 *pixels* para a direita (positivo) em destaque na imagem. Apenas os vetores que se distanciarem em 2 *pixels* (de 5 a 9 *pixels* de deslocamento para a direita) serão considerados para o cálculo dos parâmetros mostrado na seção seguinte.

Esta análise é feita apenas nas 4 regiões externas da imagem, uma vez que, na maioria dos vídeos, o objeto (que não faz parte do movimento global da imagem) se encontra em foco e está localizado na parte central do quadro de vídeo, como podemos observar na Figura 14.



Figura 14. Objetos de seqüências de vídeo presente no centro da imagem.

Os vetores finais encontrados como sendo válidos, Figura 15, são utilizados na etapa seguinte para cálculo dos parâmetros de transformação da imagem.

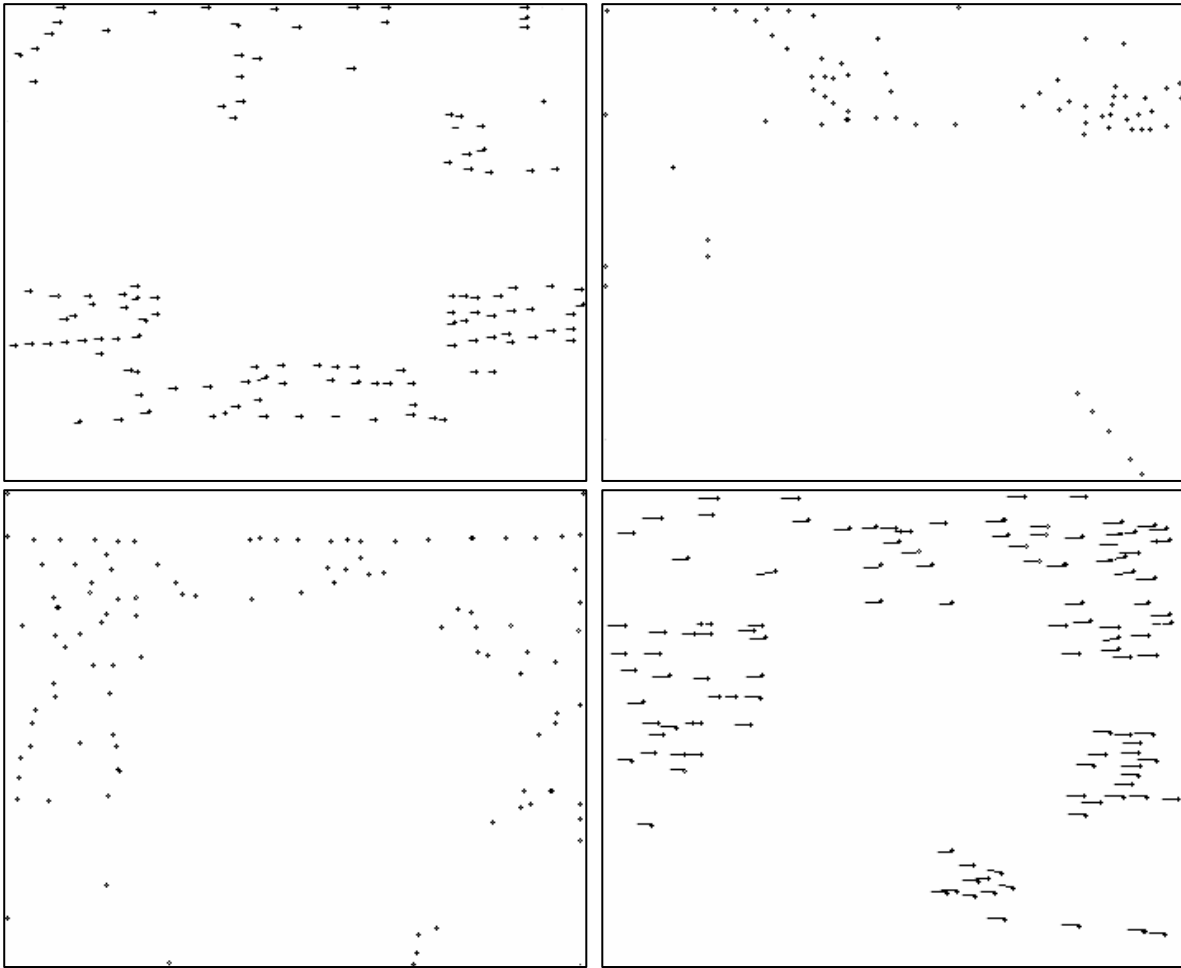


Figura 15. Vetores de movimento válidos.

3.4. Cálculo dos parâmetros de transformação

Depois de encontrados os vetores de movimento para cada ponto, monta-se um sistema de equações como feito em [19], para calcular os parâmetros de transformação entre dois quadros subseqüentes.

Através da manipulação das equações (1) e (2) obtemos as seguintes equações

$$X = x \cdot a + y \cdot b + 1 \cdot c + 0 \cdot d + 0 \cdot e + 0 \cdot f - X \cdot x \cdot g - X \cdot y \cdot h \quad (12)$$

$$Y = 0 \cdot a + 0 \cdot b + 0 \cdot c + x \cdot d + y \cdot e + 1 \cdot f - Y \cdot x \cdot g - Y \cdot y \cdot h \quad (13)$$

onde X e Y representam as coordenadas finais do vetor de movimento encontrado e x e y as coordenadas iniciais do vetor de movimento.

Representando (11) e (12) como um sistema de equações para todos os pontos encontrados obtemos

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1 \cdot X_1 & -y_1 \cdot X_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1 \cdot Y_1 & -y_1 \cdot Y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2 \cdot X_2 & -y_2 \cdot X_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2 \cdot Y_2 & -y_2 \cdot Y_2 \\ \vdots & & & \vdots & & & \vdots & \\ \vdots & & & \vdots & & & \vdots & \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_n \cdot X_n & -y_n \cdot X_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_n \cdot Y_n & -y_n \cdot Y_n \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix} = \begin{bmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \\ \vdots \\ \vdots \\ X_n \\ Y_n \end{bmatrix} \quad \text{ou} \quad A \cdot x = B \quad (14)$$

onde x_i, y_i é a posição inicial dos vetores e X_i, Y_i a posição final, n é o número de vetores de movimento encontrados.

Esta é uma metodologia eficiente, robusta e com pouco peso computacional para a detecção dos parâmetros de transformação da imagem.

Este sistema de equações é resolvido utilizando o método SVD [15] por apresentar uma boa estabilidade numérica em relação a outros métodos, por possibilitar a solução de um sistema com um número bem maior de equações que variáveis e por apresentar uma solução aproximada que minimiza a equação $e = |A \cdot x - B|$.

Para calcular x (parâmetros), precisamos de um mínimo de 4 vetores de movimento. O melhor resultado seria encontrado caso tivéssemos um vetor de movimento em cada quadrante da imagem que representasse fielmente o movimento global. O objeto em foco introduz algum erro nesta estimativa, por isso é interessante manter uma quantidade de vetores que representam o movimento global bem maior que os que representam o movimento do objeto.

No capítulo seguinte são mostrados resultados obtidos em testes feitos com algumas seqüências de vídeo.

4. Resultados

Neste capítulo serão mostrados resultados obtidos com a metodologia proposta. Primeiramente será mostrado o resultado da obtenção da resposta aos cantos de Harris nas seqüências *stefan* e *bus* para vários quadros. Em seguida serão mostrados os vetores de movimento local válidos, encontrados para estas mesmas seqüências.

Serão mostrados também os parâmetros de movimento encontrado para seqüências de vídeo padrão e para seqüências de vídeo sintetizadas. Será feita uma análise dos resultados obtidos e por fim será feita uma comparação com outros métodos existentes na literatura.

4.1. Resultado da obtenção da resposta aos cantos de Harris

A Figura 16 mostra a resposta aos cantos de Harris encontrada para a seqüência de vídeo *bus* para os quadros especificados. Analisando estas imagens, notamos que há uma grande quantidade de regiões que tem um alto valor da resposta aos cantos.

Devemos notar também que para regiões semelhantes, a resposta aos cantos também é semelhante, mesmo com alguma transformação ocorrida no quadro, o que garante que os mesmos cantos serão encontrados nos quadros.

Fica perceptível também que o detector de Harris sofre grande influência das altas freqüências contidas na imagem. Onde existir altas freqüências nas duas direções, horizontal e vertical, é caracterizado como alta chance de ser um ponto (alto valor de R).

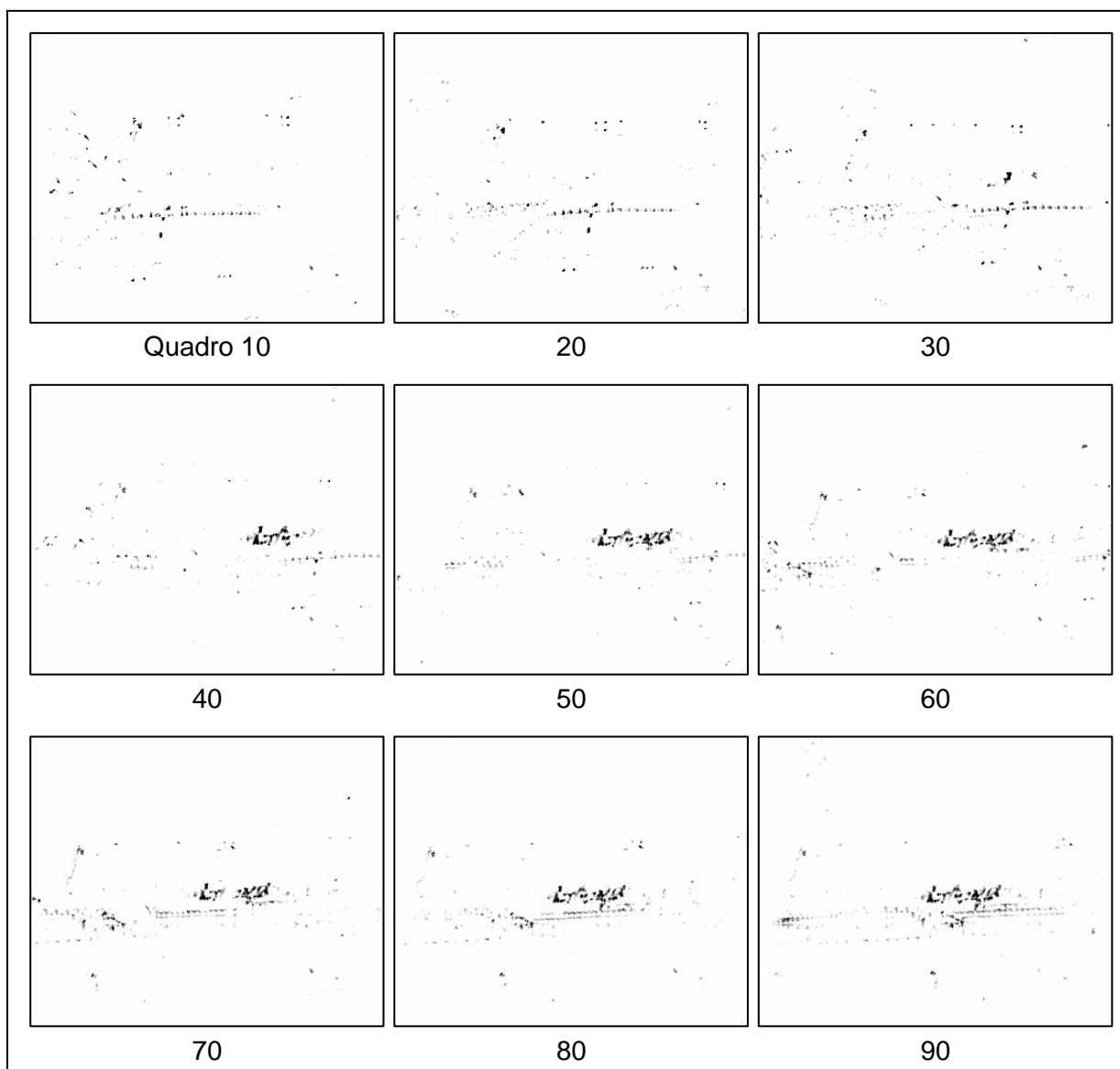


Figura 16. Resposta aos cantos de Harris da sequência *bus*.

A etapa seguinte ao cálculo da resposta aos cantos é a determinação dos pontos característicos, feito como descrito na seção 3.1. Apenas os máximos locais das imagens de resposta aos cantos são considerados efetivamente como cantos. Na Figura 17 é mostrado os pontos encontrados para os quadros apresentados anteriormente. Notamos que os mesmos cantos são encontrados nos vários quadros, independente da transformação ocorrida entre eles.

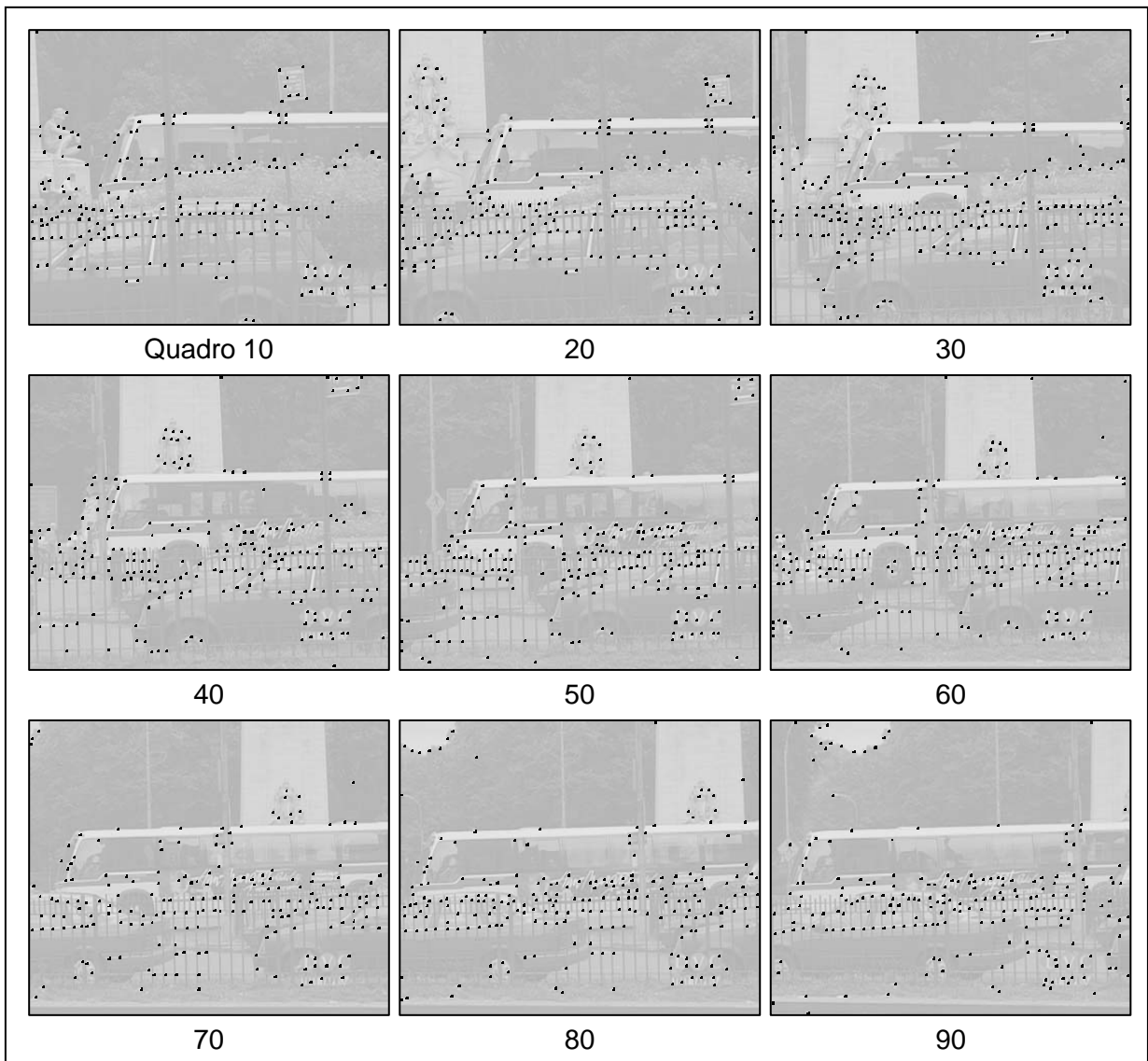


Figura 17. Pontos característicos encontrados para a seqüência *bus*.

A Figura 18 mostra a resposta aos cantos da seqüência *stefan* para os quadros apresentados. As observações são as mesmas feitas para a seqüência *bus*, apresentada anteriormente.

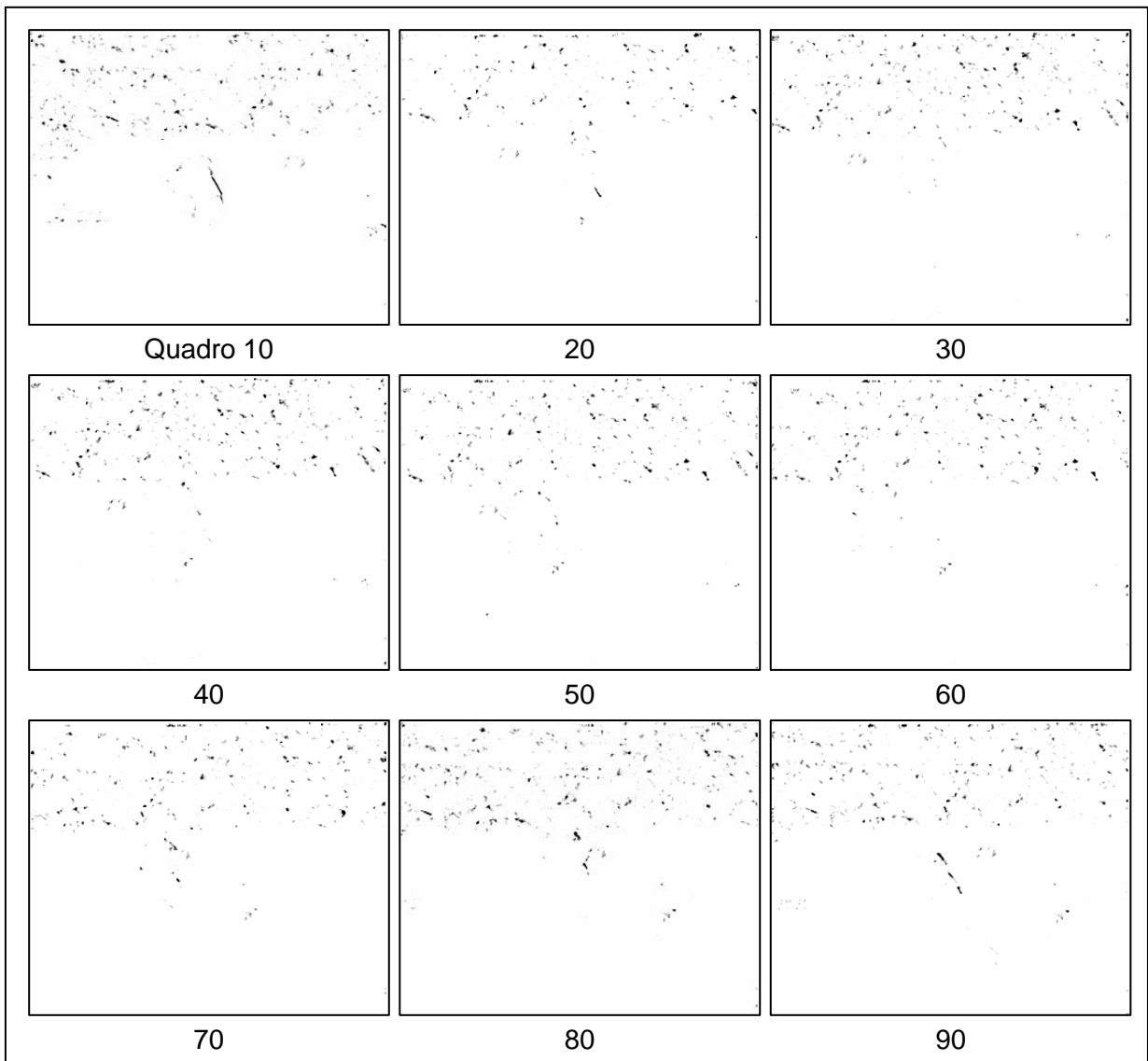


Figura 18. Resposta aos cantos de Harris da seqüência *stefan*.

A Figura 19 apresenta os pontos característicos encontrados para a seqüência *stefan*. Além dos pontos encontrados no plano de fundo da seqüência, também é mostrado que pontos são encontrados no objeto em foco, o tenista.

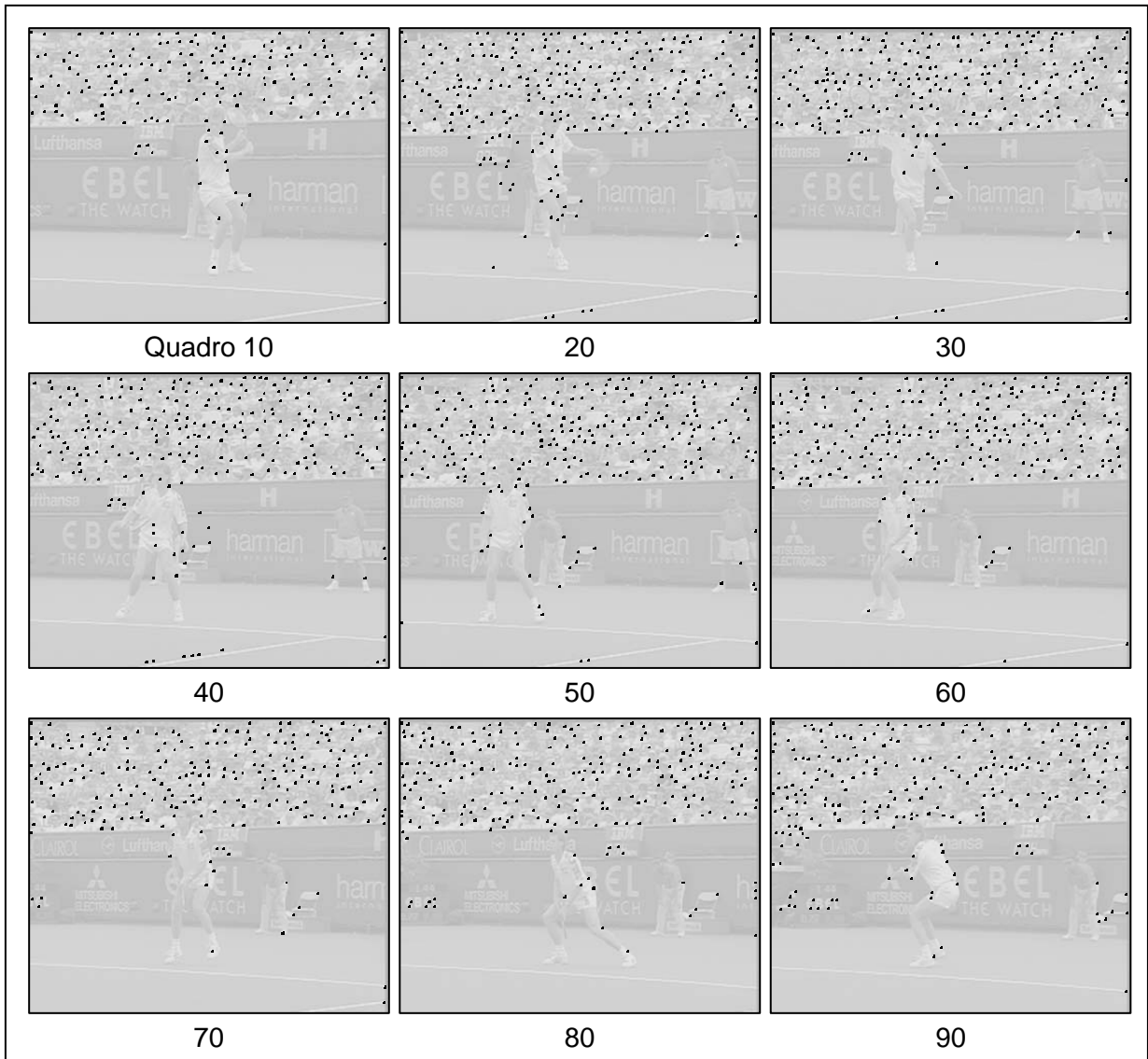


Figura 19. Pontos característicos encontrados para a seqüência *stefan*.

A quantidade de pontos característicos encontrados fica determinada pelo limiar utilizado e pela constante k da equação (8). Fazendo uma análise destes pontos, fica possível afirmar que a quantidade de pontos encontrada é menor que a quantidade de blocos que seriam estudados se fosse feita a análise BMA.

4.2. Resultados da obtenção dos vetores de movimento local válidos

Após a determinação dos pontos característicos em dois quadros subsequentes do vídeo, são obtidos os vetores de movimento local como foi descrito na seção 3.2. Na Figura 20 são mostrados os vetores de movimento encontrados para a seqüência de vídeo *bus* para os quadros informados. Observamos a predominância de movimento da câmera para a esquerda, contrário ao movimento do plano de fundo.

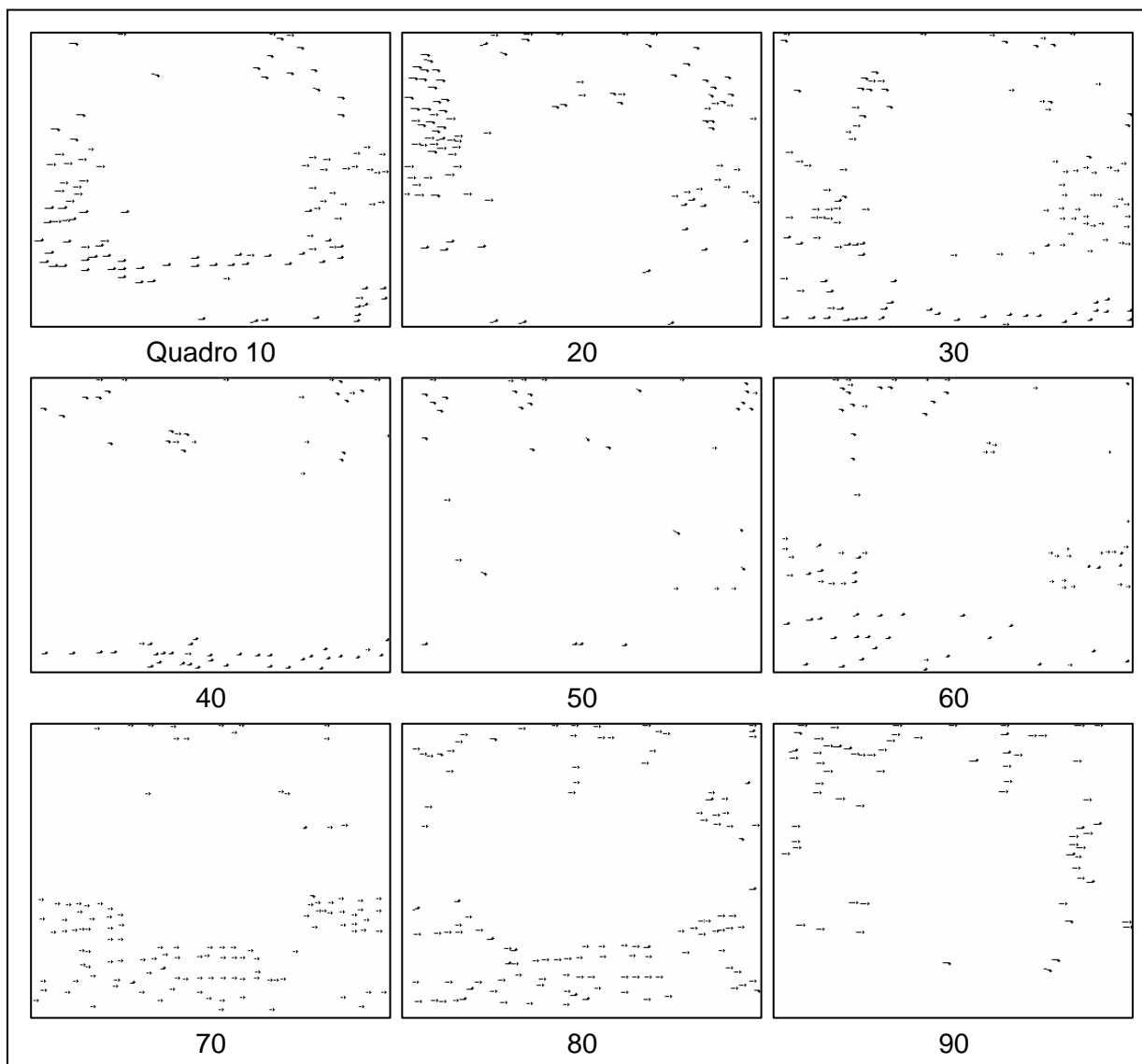


Figura 20. Vetores de deslocamento local encontrados para a seqüência *bus*.

A Figura 21 mostra os vetores encontrados para a seqüência *stefan*. Nesta seqüência são encontrados movimentos de translação, representados aqui com os quadros 10, 60, 70 e 80, movimentos de zoom com translação, quadro 20 e ainda quadros sem movimentação de câmera, quadros 30, 40, 50 e 90.

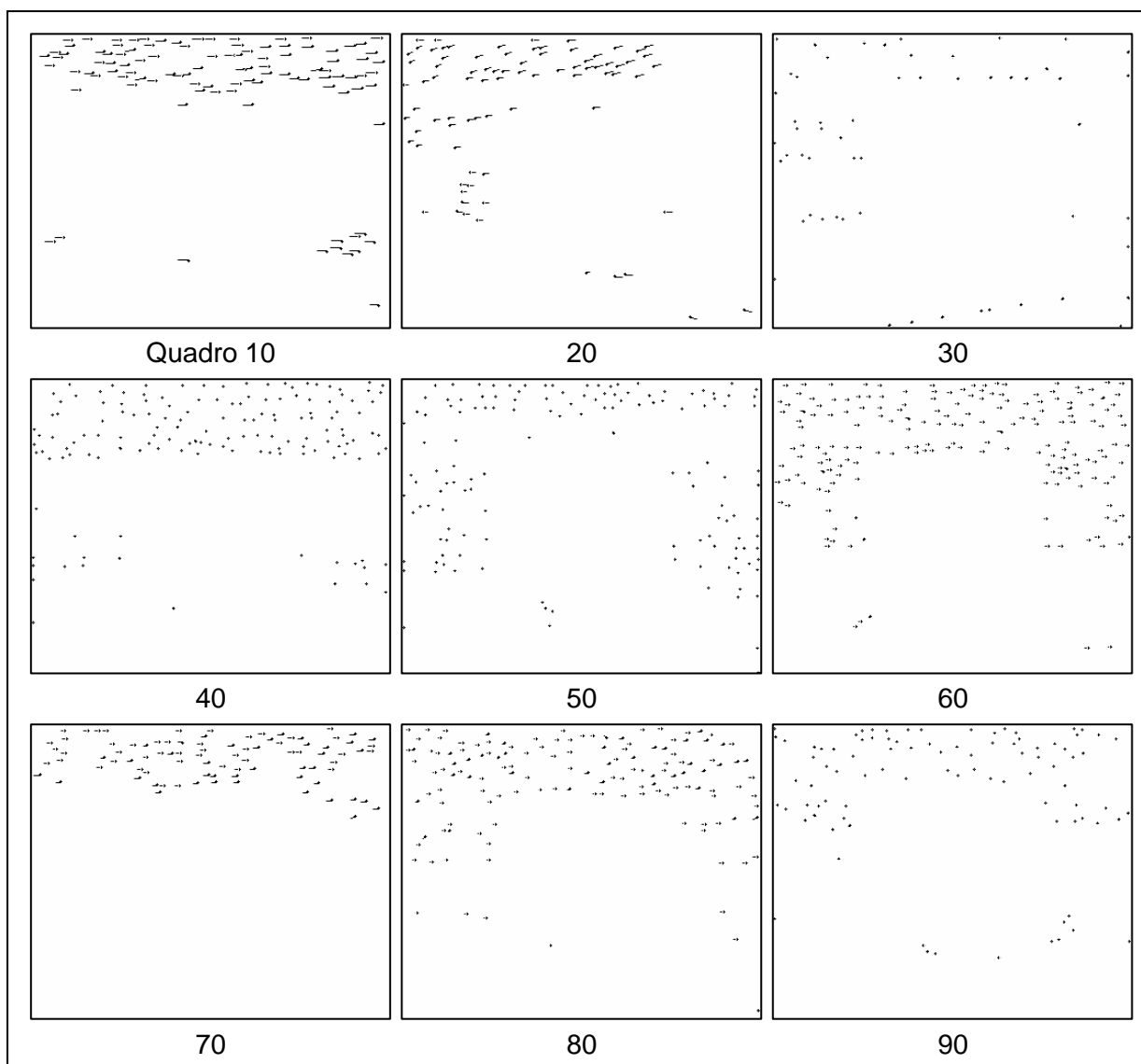


Figura 21. Vetores de deslocamento local encontrados para a seqüência *stefan*.

Notamos que os vetores encontrados são os próximos a borda da imagem. Isto é devido ao critério de validação dos vetores, onde os centrais são desconsiderados, uma vez que representariam um possível objeto.

4.3. Resultados da compensação de movimento global

Nesta seção serão mostrados a diferença entre quadros subsequentes com e sem compensação do movimento global.

A Figura 22 mostra a diferença entre quadros, $\text{abs}(I_{n-1} - I_n)$, da sequência *bus* sem compensação de movimento.

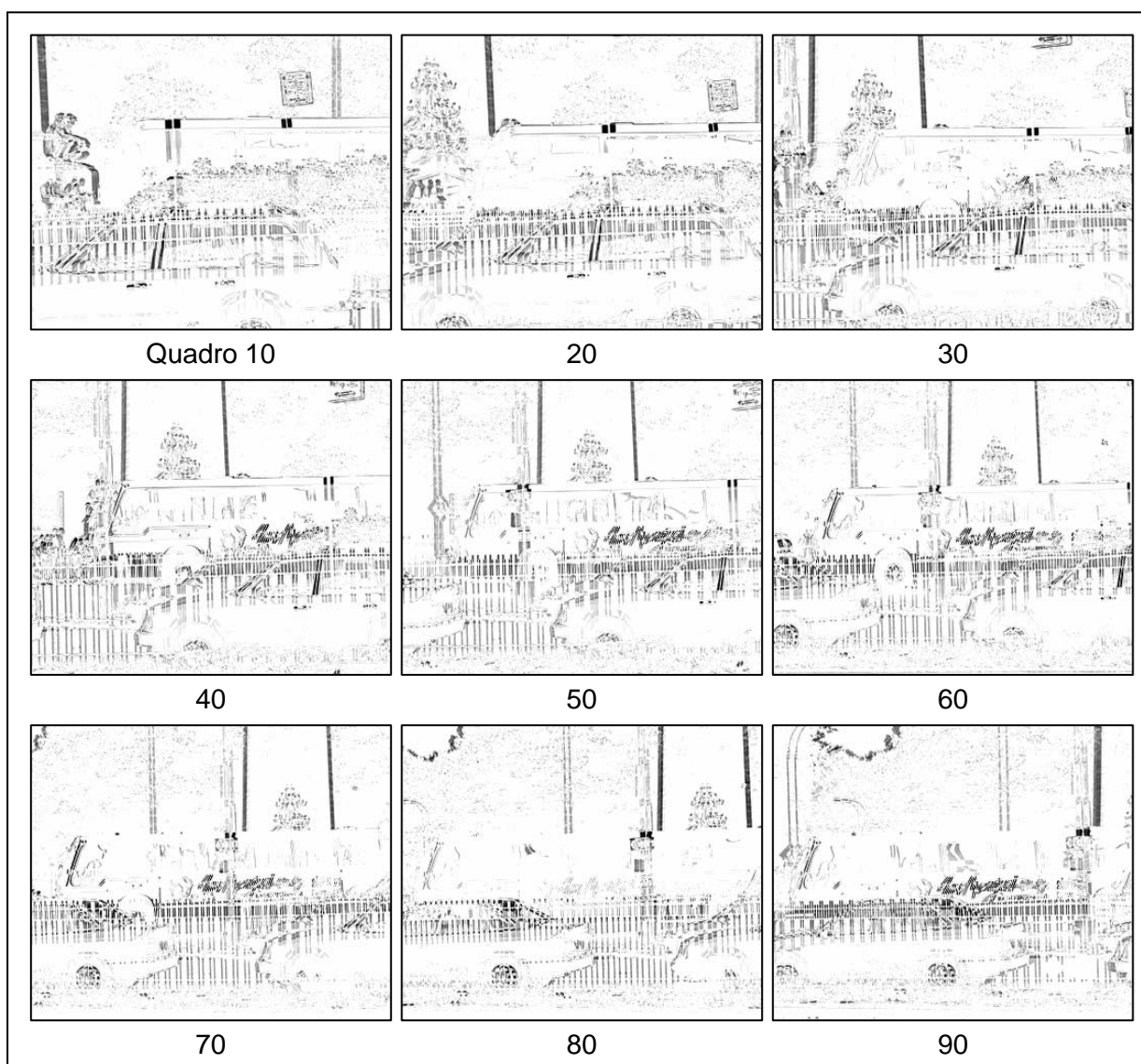


Figura 22. Diferença entre quadros da sequência *bus* sem compensação de movimento global

A Figura 23 mostra a diferença entre um quadro e o anterior compensado com a movimentação global detectada, $\text{abs}(T(I_{n-1}) - I_n)$, onde T é a transformação de acordo com as equações (1) e (2) e com os parâmetros de movimentação global detectados entre os dois quadros.

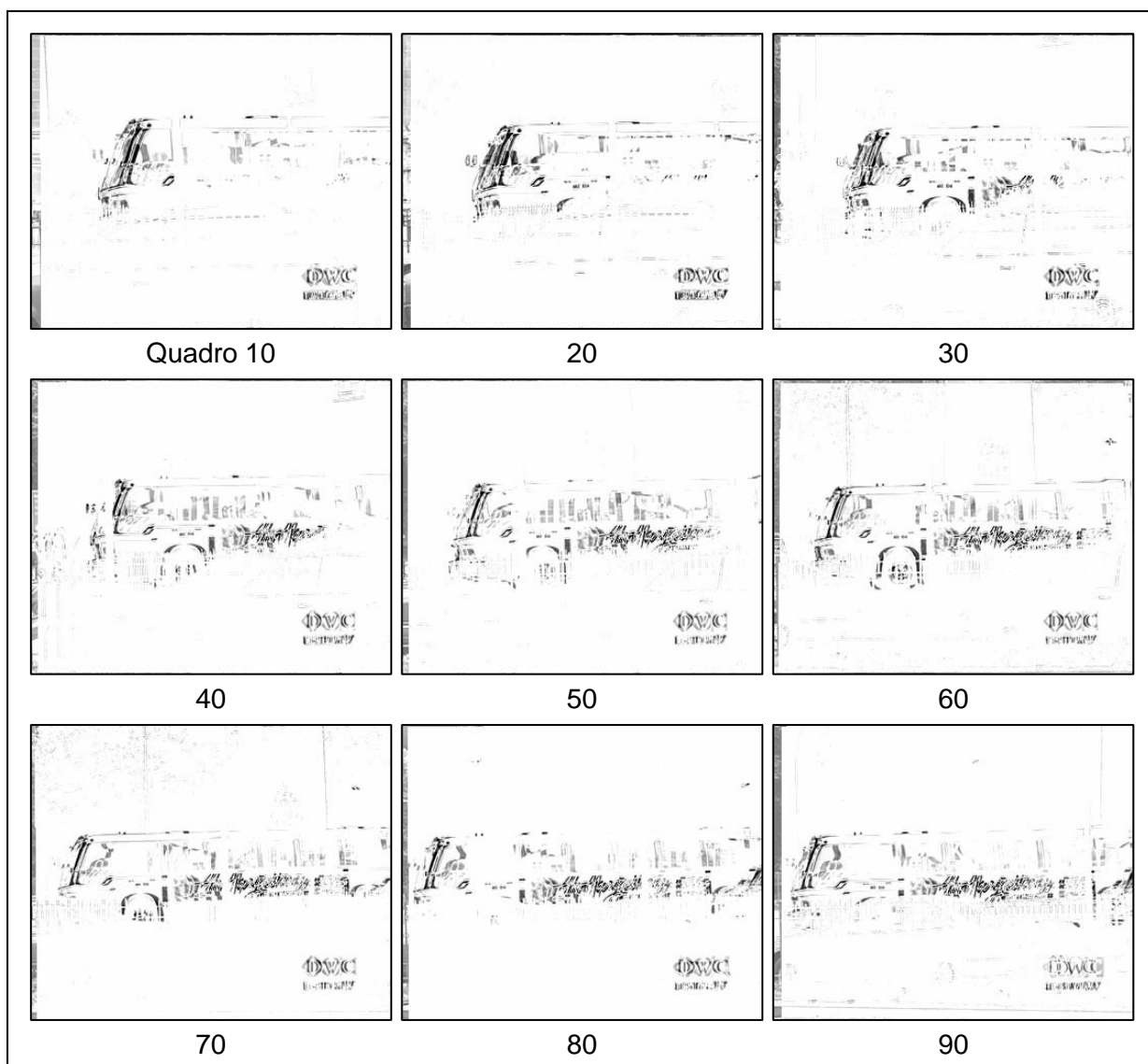


Figura 23. Diferença entre quadros da sequência *bus* com compensação de movimento global.

Podemos notar que com a compensação do movimento global, a diferença entre o quadro atual e o anterior transformado diminui muito. Também fica visível o destaque do objeto em foco, nesta seqüência, o ônibus, juntamente com outros objetos que se encontram na cena, como por exemplo, o anuncio da emissora. Esta característica pode ser utilizada para a segmentação (separação) dos objetos contidos no vídeo.

A Figura 24 apresenta a diferença entre dois quadros seguintes, da seqüência *stefan*. Do mesmo modo como mostrado para a seqüência *bus*, também é feita a diferença sem compensação de movimento global.

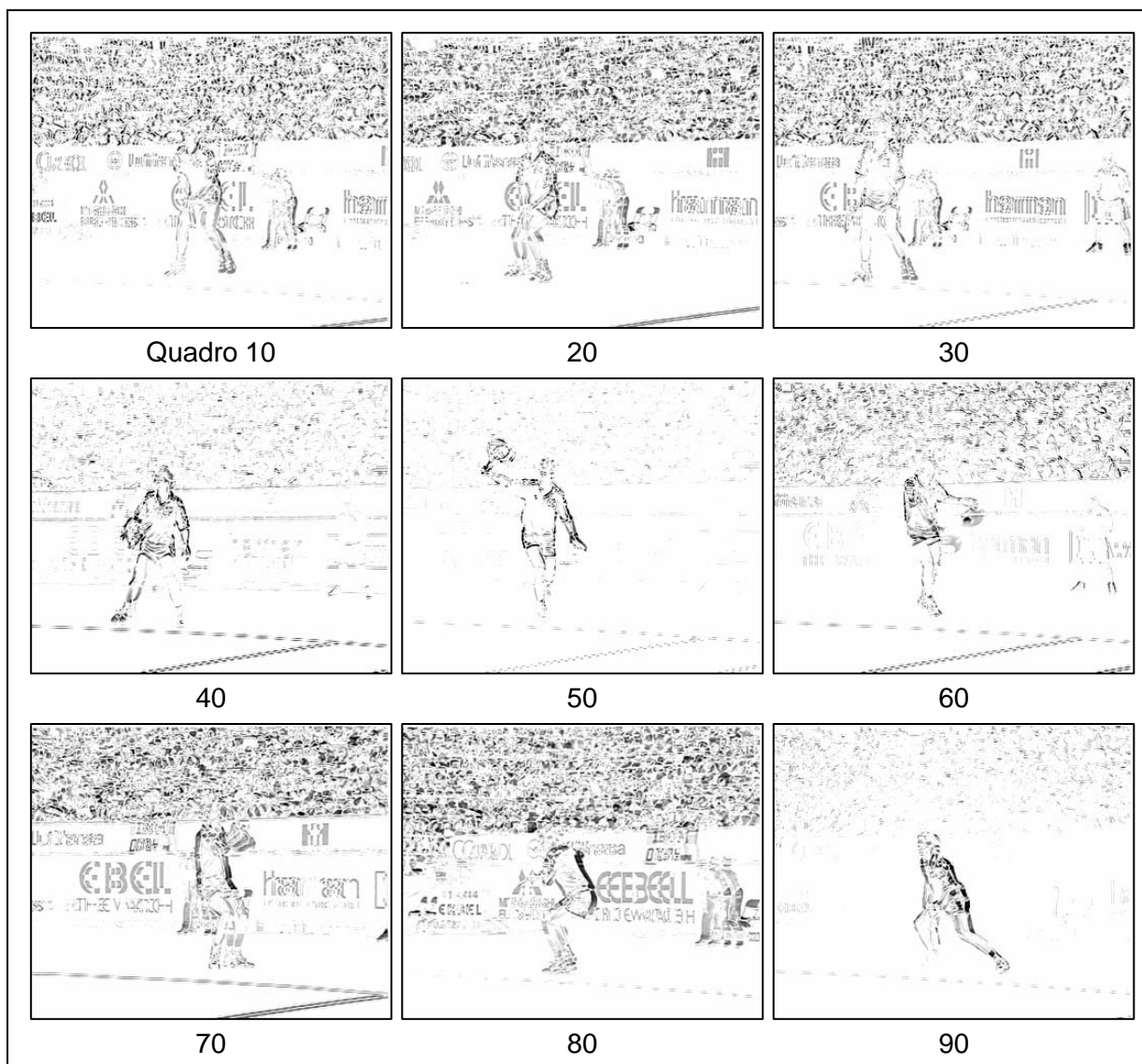


Figura 24. Diferença entre quadros da sequência *stefan* sem compensação de movimento global

Agora, quando utilizamos a transformação da imagem de acordo com os parâmetros de movimentação global encontrados, fica perceptível o casamento entre os dois quadros subsequentes.

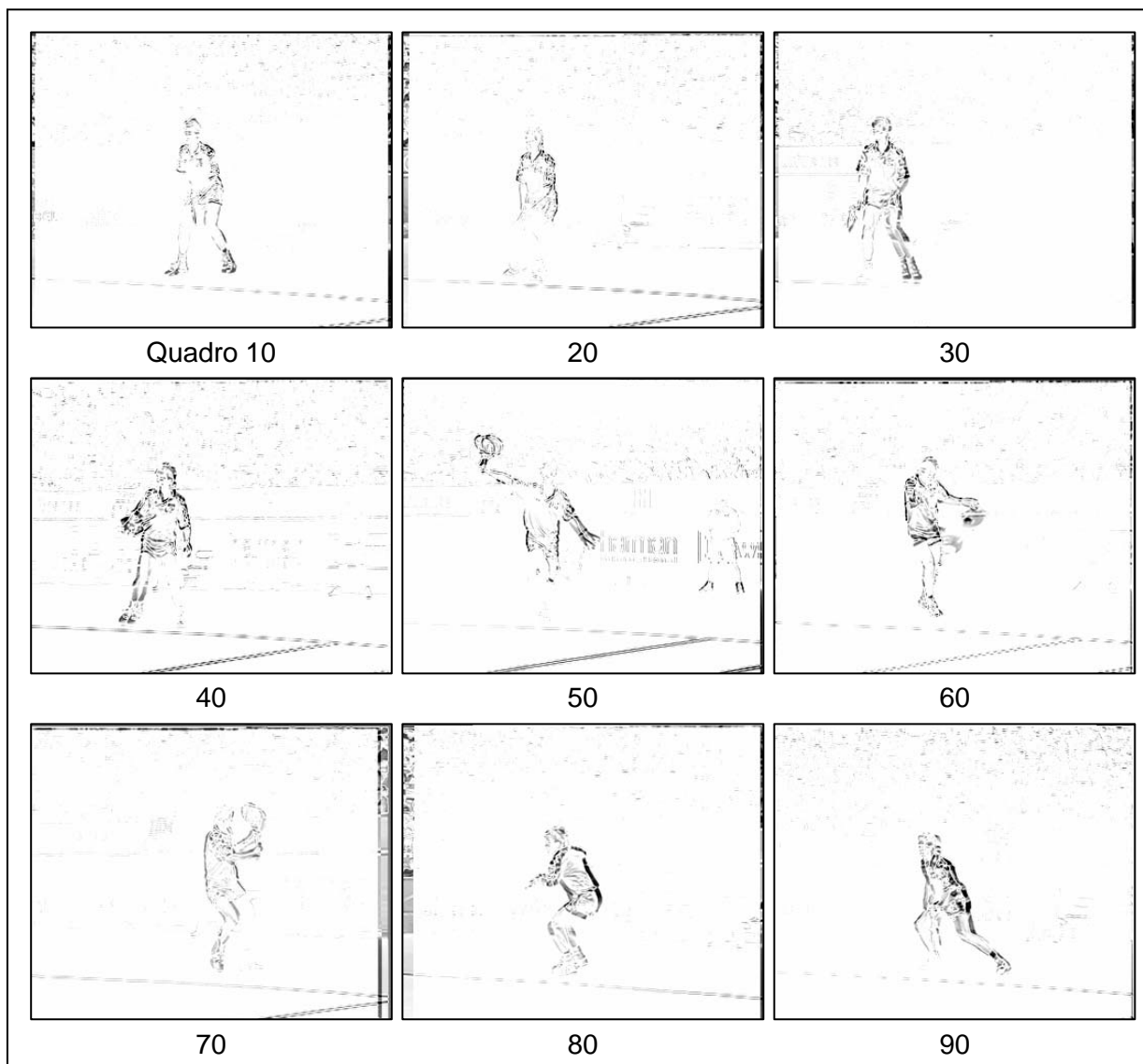


Figura 25. Diferença entre quadros da sequência *stefan* com compensação de movimento global.

Podemos tirar algumas conclusões com estas imagens apresentadas. Primeiramente, notamos que, a quantidade de dados necessária

para a transmissão do quadro, que foi compensado pelo movimento global detectado, diminui bastante. A compensação do movimento provoca um casamento da maior parte dos quadros, necessitando assim, transmitir somente a variação entre um quadro e outro, ou seja, a diferença entre eles.

Uma outra conclusão a que se pode chegar é que, com a compensação do movimento global do vídeo, fica mais fácil a detecção dos objetos contidos na imagem e por consequência a extração destes objetos, visto que, é um dos objetivos das novas metodologias de codificação de vídeo como o MPEG-4.

Fica possível observar também que a determinação da movimentação da câmera, ou seja, o movimento global é bastante precisa, que é o objetivo principal deste trabalho.

4.4. Resultado da obtenção dos parâmetros de movimento global

Nesta seção são mostrados os parâmetros obtidos para seqüências de vídeo geradas com movimentos de câmera pré-determinados e também para algumas seqüências de vídeo utilizadas como padrão na literatura.

4.4.1. Seqüências sintetizadas

Com o intuito de demonstrar a precisão e qualidade da estimação dos parâmetros de transformação, foram gerados alguns vídeos com algumas transformações pré-determinadas como translação, rotação, zoom, inclinação.

Estas seqüências foram feitas a partir de imagens fotográficas para representar o mais fiel possível uma aquisição de vídeo comum. A Figura 26 mostra as imagens utilizadas para geração destes vídeos e a representação de três quadros gerados com as transformações determinadas.

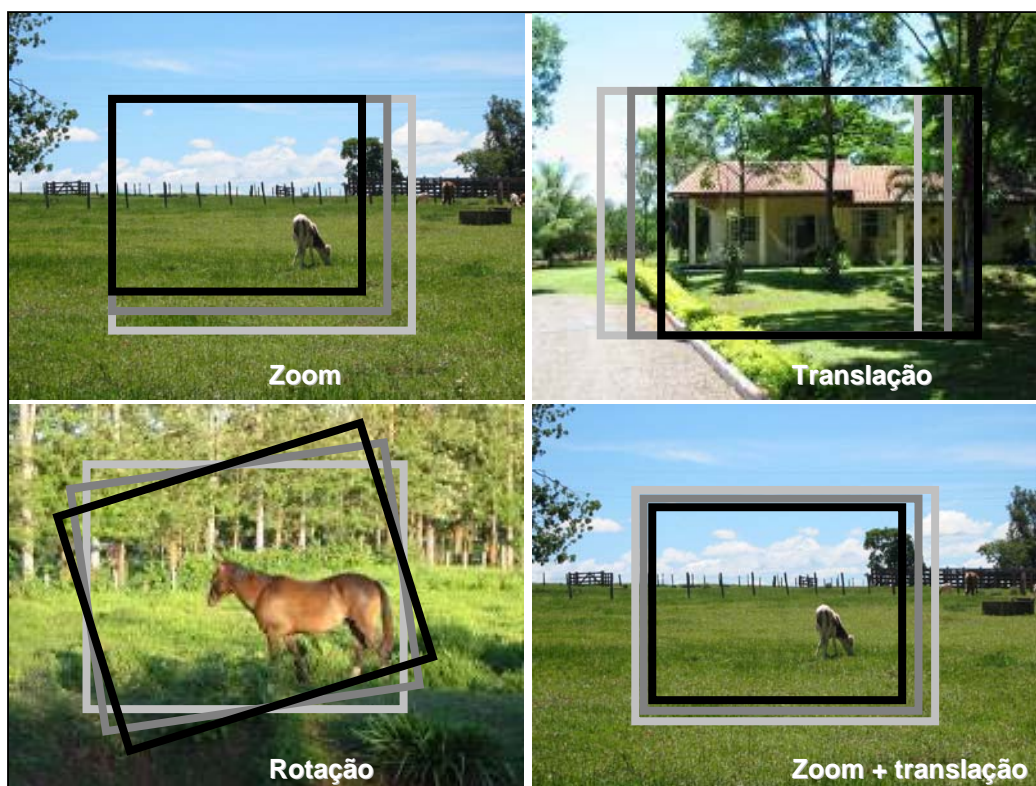


Figura 26. Vídeos gerados com transformações determinadas.

Estas seqüências geradas não contem objetos em movimento devido à dificuldade de geração deste tipo de vídeo. Mas para uma validação da metodologia, estes são muito úteis.

Depois de gerados os vídeos, todos com 24 quadros, são obtidos os parâmetros e comparados com os originais. Os resultados são mostrados a seguir.

- Seqüência nº 1:

Esta seqüência foi gerada com o deslocamento em 1 *pixel* para a direita do quadro em relação ao quadro anterior, ou seja, ao final de 24 quadros, o ultimo estaria a uma distância de 24 *pixels* para a direita do primeiro quadro.

A Figura 27 mostra a evolução do parâmetro de translação horizontal (parâmetro c) da sequência gerada e também do detectado pela metodologia proposta.

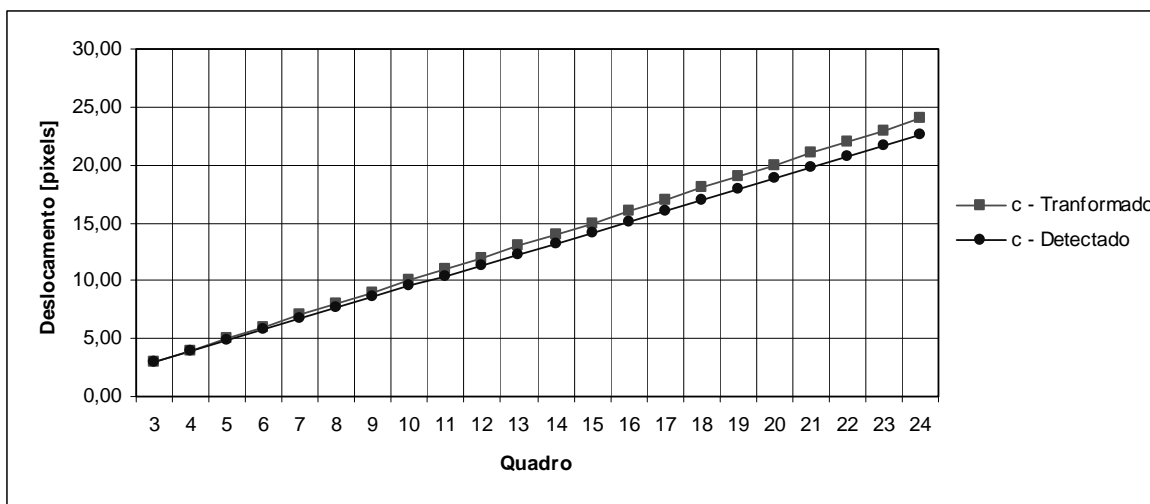


Figura 27. Translação horizontal de 1 *pixel*/ para a direita.

Os demais parâmetros não sofreram alterações em relação ao primeiro quadro.

- Sequência nº 2:

Esta sequência foi gerada do mesmo modo que a anterior, só que agora com o deslocamento para a esquerda.

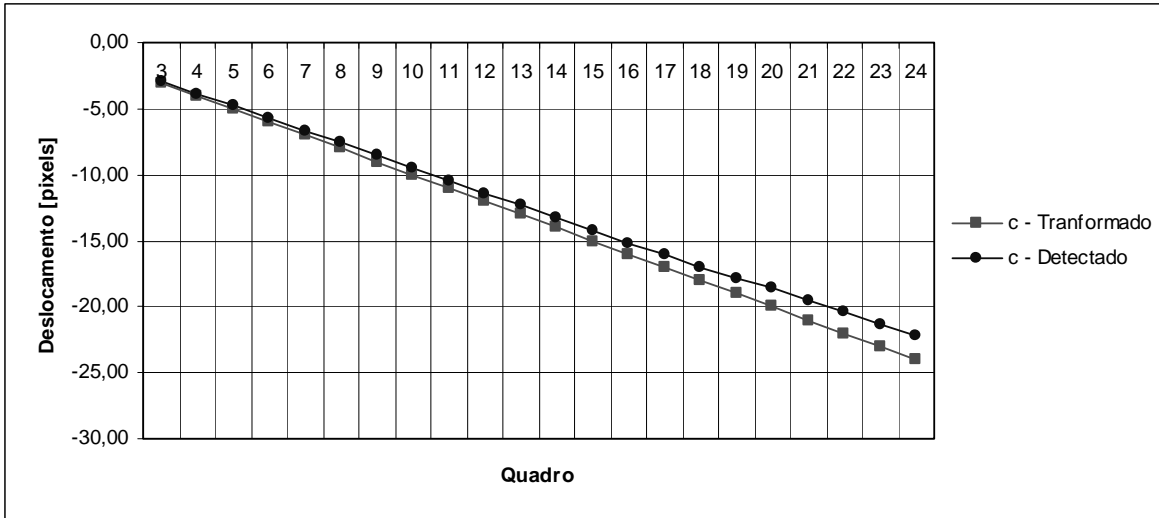


Figura 28. Translação horizontal de 1 *pixel* para a esquerda.
Os demais parâmetros não sofreram nenhuma alteração.

- Sequência nº 3:

Esta sequência foi gerada com deslocamento de 1 pixel para cima do quadro em relação ao quadro anterior.

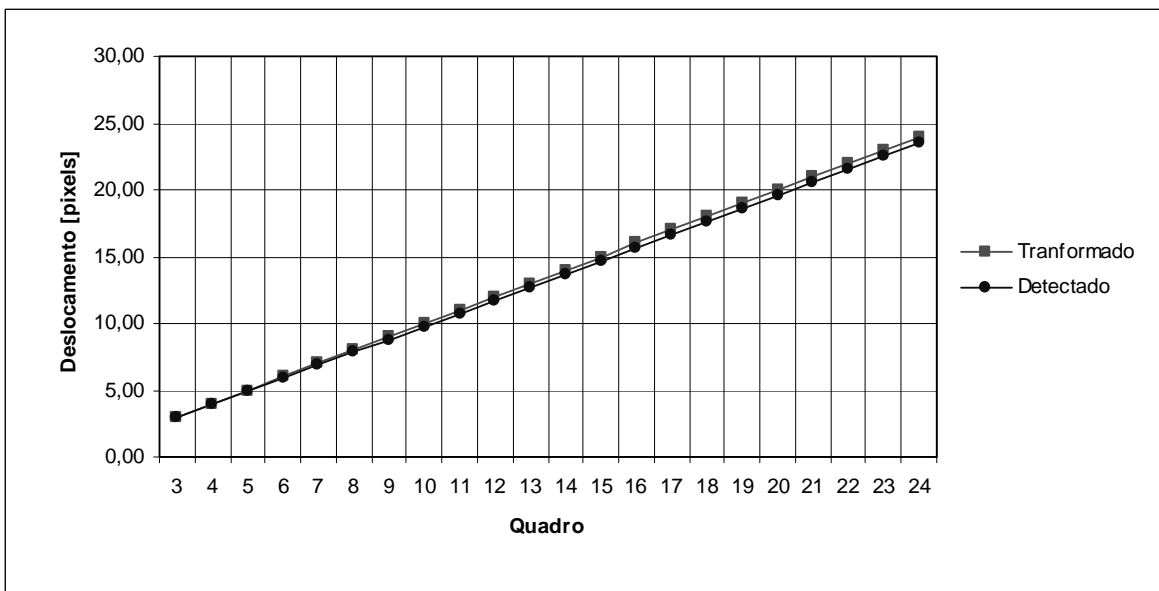


Figura 29. Translação horizontal de 1 *pixel* para cima.

Nenhum outro parâmetro sofreu alguma alteração considerável.

- Seqüência nº 4:

Esta seqüência foi gerada do mesmo modo que a primeira, porém com deslocamento de 3 *pixels*. No final dos 24 quadros, o ultimo estará a uma distância de 72 *pixels* para a direita do primeiro quadro. A Figura 30 mostra a evolução dos parâmetros da seqüência nº 4 gerada.

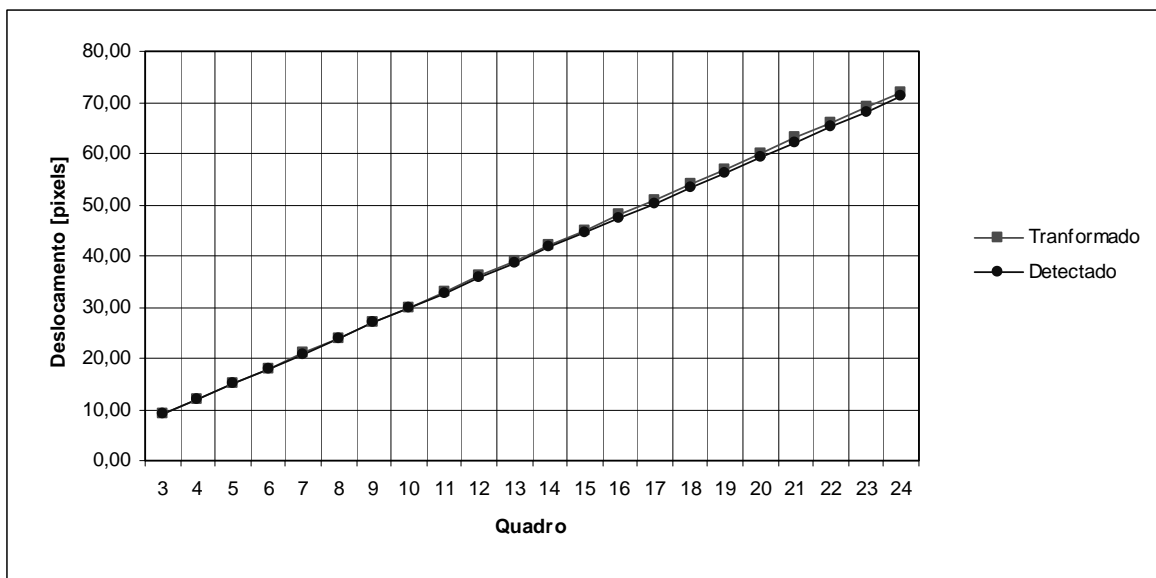


Figura 30. Translação horizontal de 3 *pixel* para a direita.

Podemos notar a melhor estimativa do parâmetro de translação. Também, como nas anteriores, os outros parâmetros não sofreram alterações significativas.

- Seqüência nº 5:

A seqüência de numero 5 foi gerada para a rotação da imagem. Os parâmetros de escala (parâmetros *a* e *e*) sofrem propagação de 0,01 para cima em cada quadro terminando em 1,26 no ultimo quadro. Os parâmetros de

inclinação (b e d) sofrem alterações de 0,01 e -0,01 respectivamente, em cada quadro o que resultaria em 0,24 para o ultimo quadro em relação ao primeiro. A translação e os outros parâmetros não são utilizados. A composição destes parâmetros provoca a rotação da imagem em cada quadro.

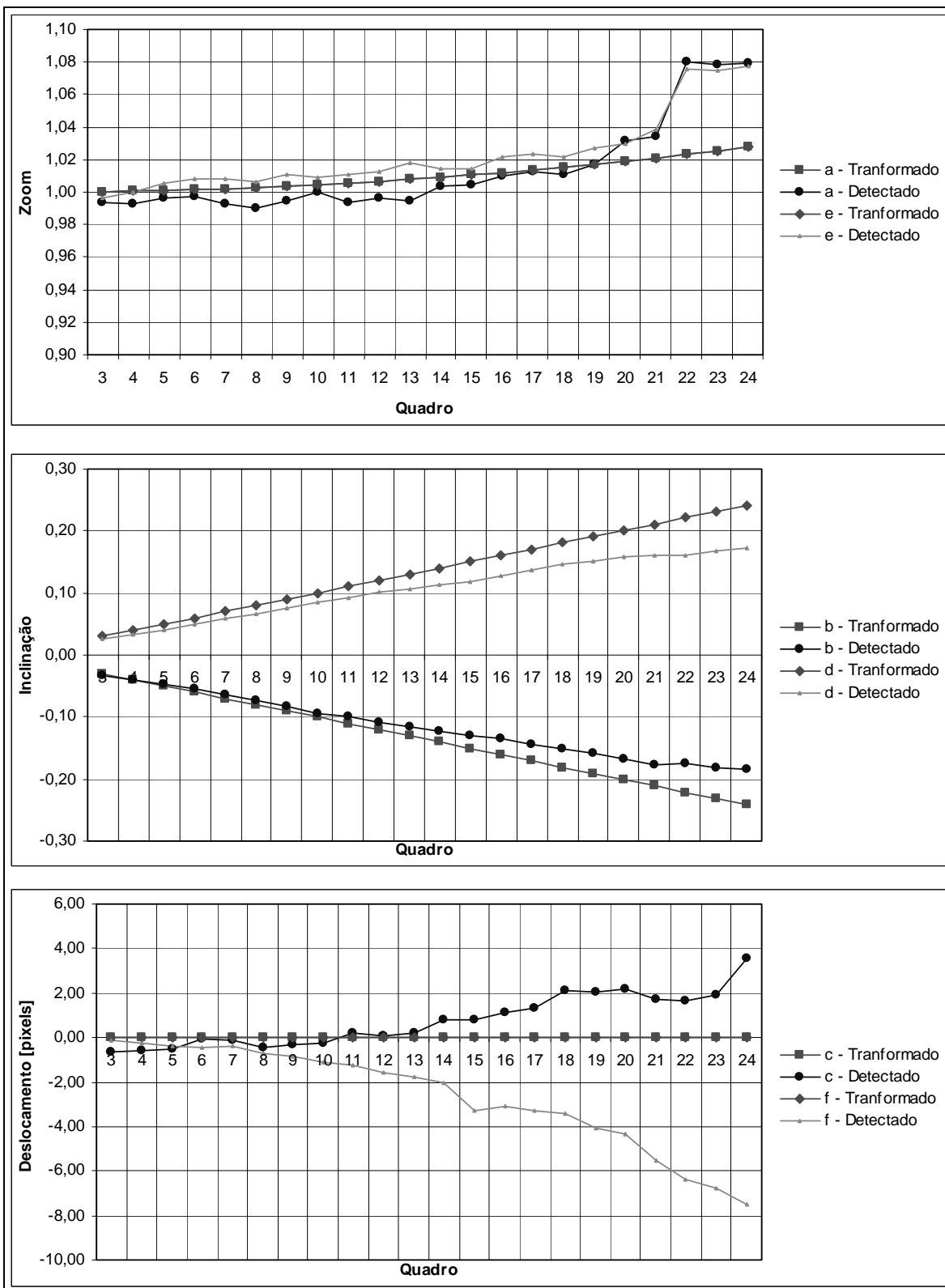


Figura 31. Rotação anti-horária.

Verificamos que os parâmetros encontrados são diferentes daqueles utilizados na geração do vídeo. Isto não determina que a detecção é errônea, pois se for analisada a diferença entre os quadros, é percebido que ocorre o casamento entre eles. Isto indica que a transformação utilizada na análise (equações (1) e (2)) não é unívoca, ou seja, com parâmetros diferentes é possível obter os mesmos resultados. O método SVD que resolve o sistema ou $A \cdot x = B$ (14) encontra uma solução que minimiza o critério de erro, porém não é necessariamente a solução utilizada na geração do vídeo.

- Seqüência nº. 6:

Esta seqüência e a próxima foi gerada para demonstrar a ocorrência de zoom nos vídeos. Foi utilizada a evolução de -0,01 para os parâmetros de escala, resultando no final de 24 quadros em 0,78 para os parâmetros a e e . A Figura 32 mostra a evolução para estes parâmetros.

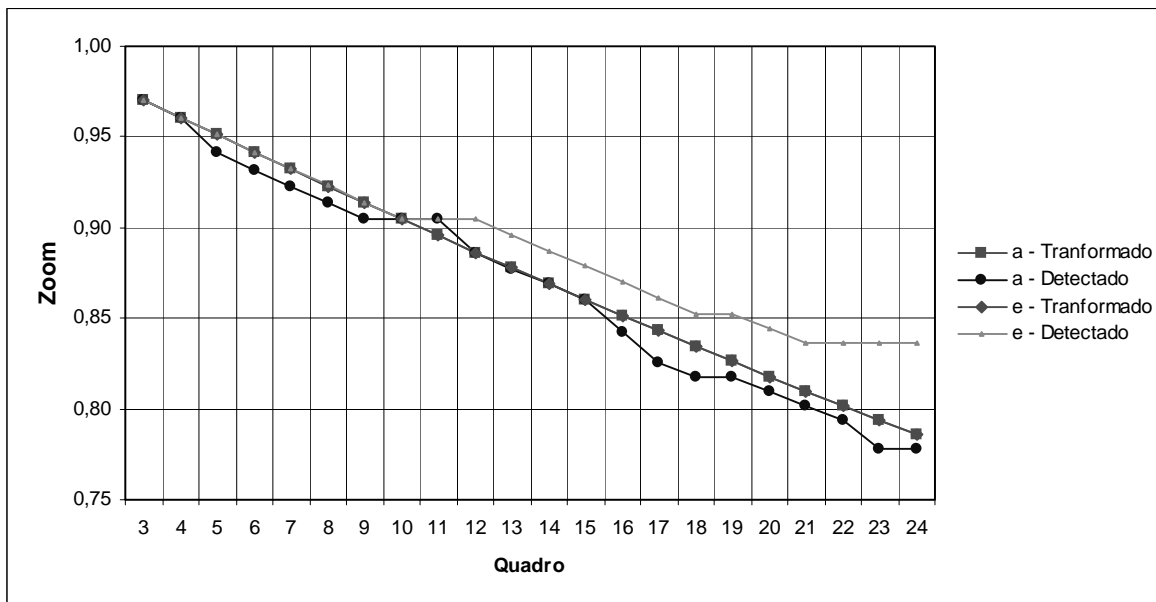


Figura 32. Zoom In.

Verificamos que é encontrada a tendência de evolução dos parâmetros. Porém é observado um erro no final da seqüência.

- Seqüência nº 7:

Esta seqüência foi gerada da mesma forma que a anterior, porém com a evolução positiva do parâmetro de escala. Os parâmetros são apresentados na Figura 33.

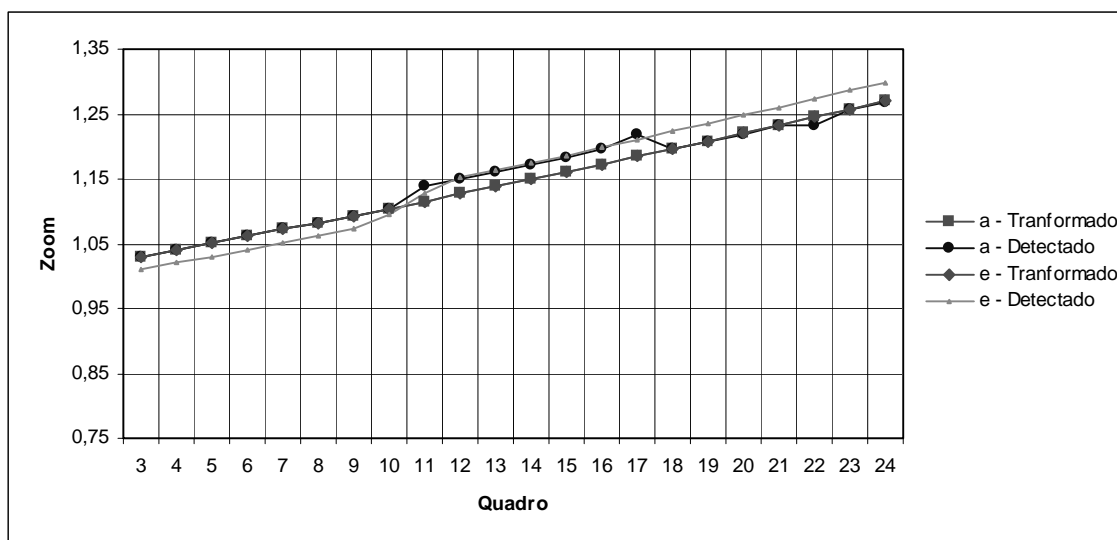


Figura 33. Zoom Out.

Para esta seqüência, os parâmetros detectados são mais próximos dos reais.

- Seqüência nº 8:

A seqüência 8 tem o objetivo de mostrar o zoom central na imagem, um evento comum nas seqüências de vídeo. Para representar esta movimentação de câmera, são utilizados os parâmetros de translação juntamente com os parâmetros de escala. Os parâmetros de escala são evoluídos em 0,01 e os

parâmetros de translação são alterados em 1 *pixel*. O resultado é a ocorrência de zoom próximo ao centro da imagem.

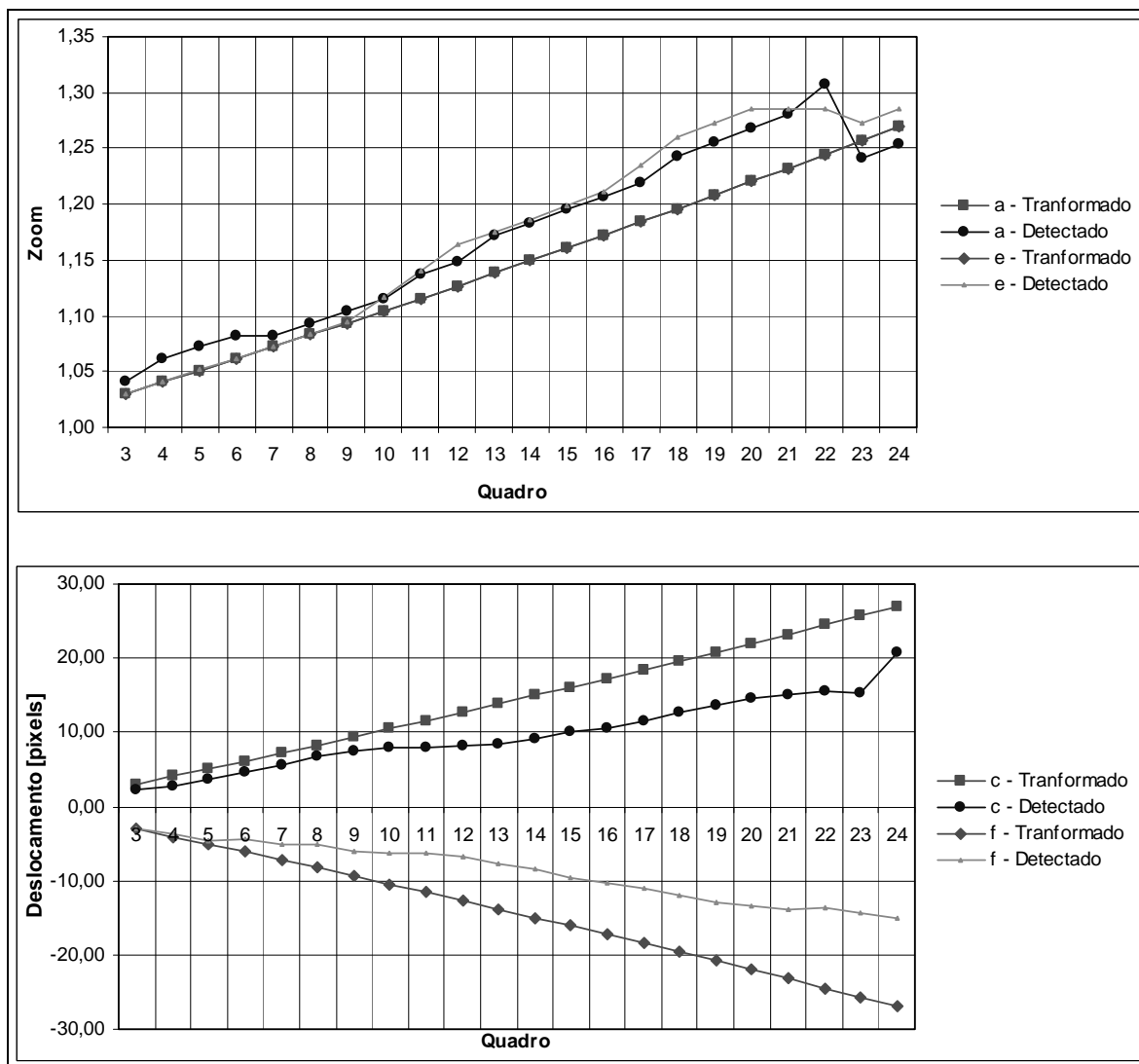


Figura 34. Zoom com translação.

Da mesma forma como ocorre na seqüência de número 5, os parâmetros detectados diferem dos utilizados para gerar a seqüência, porém, não que dizer que o resultado seja errôneo.

Se não fizermos a etapa da separação da imagem em blocos, para validação dos vetores (seção 3.3), fica prejudicada a detecção dos parâmetros para os vídeos com zoom, pois a grande maioria dos vetores encontrados aponta para o centro ou para fora da imagem, ou seja, todos apontam para direções

diferentes. Então, a média dos vetores, ou a moda dos vetores vai ser zero ou próxima de zero. Também, a estimação dos parâmetros seria contaminada pelos vetores correspondentes aos objetos da imagem ocasionando erros na estimação.

Verificando os resultados obtidos para seqüências padrões é possível afirmar que a metodologia proposta atende as necessidades e que os resultados obtidos são próximos dos esperados.

4.4.2. Seqüências padrão

As seqüências padrões utilizadas neste estudo foram as seqüências *stefan*, *hall*, *silent* e *bus*, por apresentarem várias características encontradas em qualquer seqüência de vídeo, bem como transformações importantes que validam a metodologia proposta. São encontrados nestas seqüências etapas com zoom, translação horizontal, inclinação, imagens estáticas com objetos focados no centro e sem objetos, entre outros movimentos importantes.

Na Figura 35 são mostrados os resultados dos parâmetros de zoom (a e e) e translação (c e f) para a seqüência *stefan*. O parâmetro a determina a transformação de escala horizontal (x), o parâmetro e , a transformação de escala vertical (y), o parâmetro c , o deslocamento horizontal (dx) e o parâmetro f , o deslocamento vertical (dy).

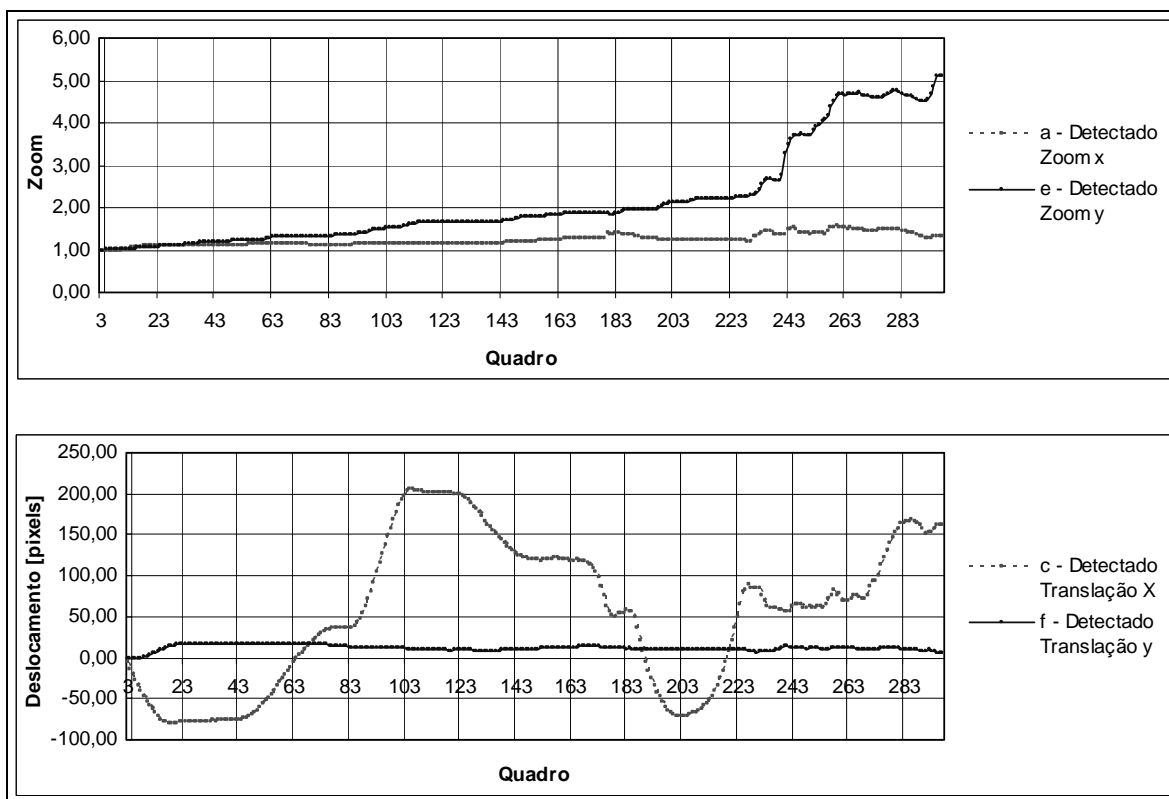


Figura 35. Evolução dos parâmetros de transformação da seqüência *stefan*.

Foi verificado que a seqüência *stefan* entre os quadros 225 e 265 ocorrem movimentos de translação maiores que 16 *pixels*. Isto provoca um erro na determinação dos parâmetros. Na figura anterior, os valores mostrados entre estes quadros são incorretos. Isto fica verificado ainda comparando-se com resultados de outros autores o que será feito na seção seguinte.

Na Figura 36 mostra os resultados para os parâmetros de escala e translação para a seqüência *bus*.

Os parâmetros *b* e *d* têm influência significativa a partir da segunda casa decimal. Porém foi notado que nos vídeos estudados ocorrem apenas variações da ordem de 0,01 a 0,03 para estes parâmetros. Apesar da variação pequena, elas são importantes para as transformações de inclinação e *pan* das imagens. Do mesmo modo, os parâmetros *g* e *h* têm influência significativa na quarta casa decimal. Estes parâmetros são mantidos na análise devido à

cobertura de transformações comuns. Devido a estas pequenas variações, estes parâmetros não são mostrados.

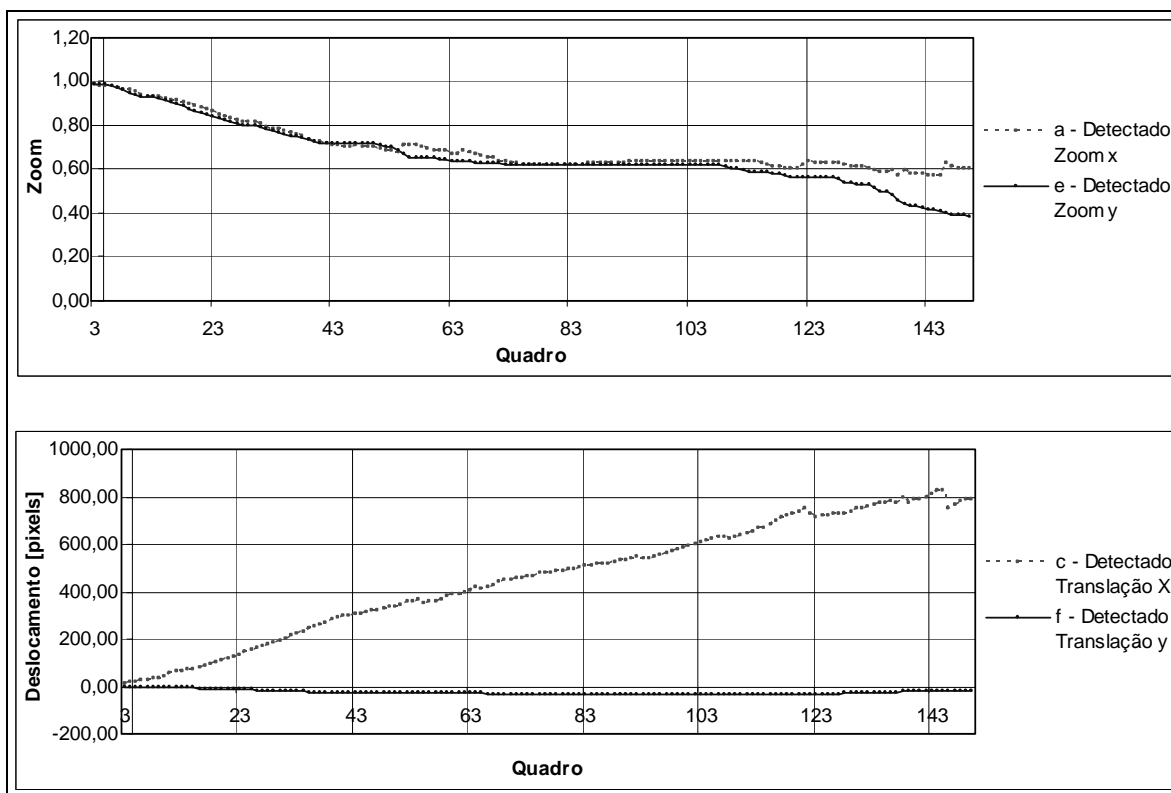


Figura 36. Evolução dos parâmetros de transformação da sequência *bus*.

Na Figura 37 são mostrados os resultados para os parâmetros da sequência *hall*.

Apesar do vídeo ser uma sequência sem movimentação de câmera, é possível notar que a metodologia encontra alguns movimentos de translação e também de zoom. Porém se observado que em 280 quadros seqüentes, houve apenas um deslocamento de 2 *pixels* na horizontal e 1 *pixel* na vertical.

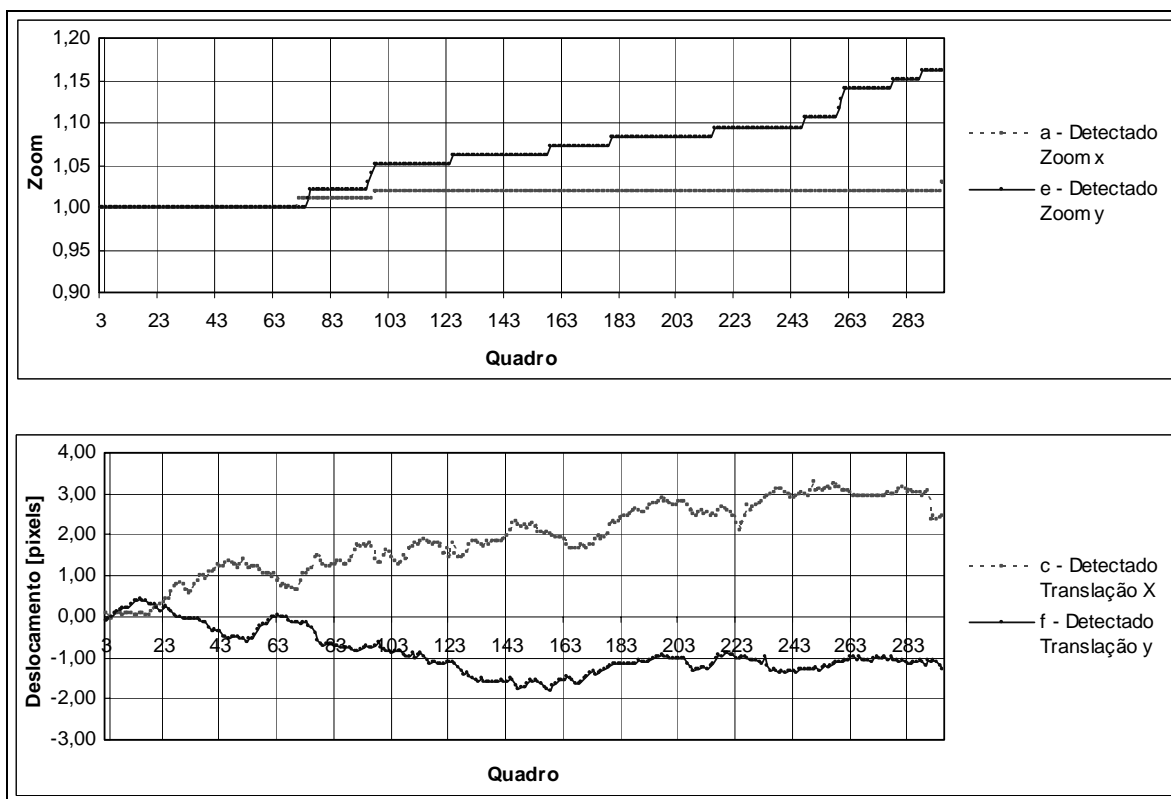


Figura 37. Evolução dos parâmetros de transformação da sequência *hall*.

É possível afirmar que a pequena influência dos objetos, que não se encontram no centro da imagem, provoca algum erro na medição, porém não inviabiliza a utilização da técnica. Talvez uma solução que não provocaria estes erros era a melhor separação do objeto.

A Figura 38 mostra a evolução dos parâmetros para a sequência *silent*.

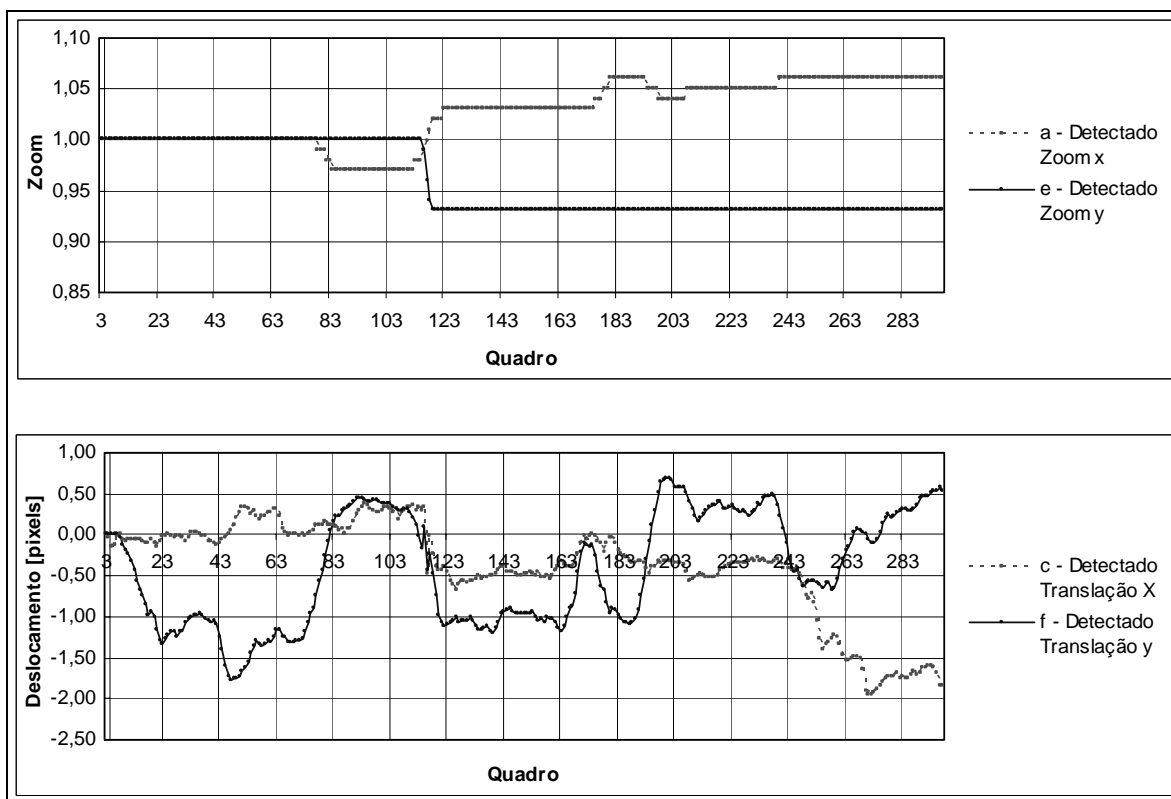


Figura 38. Evolução dos parâmetros de transformação da seqüência *silent*.

Da mesma forma, como ocorrido na seqüência *bus* a movimentação detectada é pequena, pois também não há movimento de câmera deve-se a erros introduzidos pelo objeto que parte dele interfere nas regiões de análise.

4.5. Análise dos resultados

Na seção anterior foram mostrados os resultados obtidos para seqüências padrões e para seqüências sintetizadas.

Para as seqüências geradas com movimentações pré-determinadas, a estimação dos parâmetros teve um bom resultado. É possível afirmar devido aos vídeos com movimentação pura, somente translação ou somente zoom, em que se forem verificados os erros da estimação dos parâmetros para o ultimo quadro, é notado que estes erros são pequenos. São calculados os erros da seguinte

forma $E = \frac{p - p'}{p} \cdot 100$, onde E é o erro encontrado, p é o valor esperado para o

parâmetro e p' o valor detectado. Foi encontrado para a seqüência 1 o erro de 5,8% para o parâmetro estimado no ultimo quadro da seqüência. Na seqüência 2 este erro foi de 7,3%. Para a seqüência 3, o erro é de 2,0%. E para a seqüência 4 no valor de 1,2%.

Podemos verificar que para a seqüência de número 4, o erro estimado para o parâmetro no último quadro é bem mais baixo que das demais seqüências. Isto é devido ao movimento do vídeo ser maior. Como a movimentação mínima dos pontos característicos detectada é de 1 *pixel*, quando se é analisado para seqüências que tem movimentação próxima de 1 *pixel*, erros da posição dos pontos e conseqüentemente, dos vetores locais encontrados são propagados gerando um acúmulo pequeno, porém significativo para o último quadro da seqüência.

Quando a movimentação é maior que 1 *pixel*, estes erros devido à imprecisão dos vetores passam a ser menor. Uma possível solução para evitar esta propagação de erros é diminuir a quantidade de quadros consecutivos analisados, assim, terminando-se a análise destes quadros, uma nova análise é começada.

Agora, analisando as seqüências geradas com rotação, zoom e zoom com deslocamento, podemos notar que a diferença no quadro 24 é bastante significativa. Isto não significa que não houve uma boa detecção do movimento global, pois os erros entre o quadro estimado e o quadro gerado são pequenos (quadros semelhantes), mas sim que existe outra solução que não é a utilizada na geração do vídeo.

Isto é devido a não-unicidade da equação de transformação que representa a movimentação global, equações (1) e (2), onde mais de um conjunto de parâmetros representa a mesma transformação. A solução encontrada pelo método SVD contempla o objetivo.

Para as seqüências padrões *stefan*, *bus*, *silent* e *hall*, temos os resultados bastante coerentes. Podemos notar apenas para a seqüência *stefan*

que entre os quadros 225 e 265 a movimentação da câmera é maior que 16 *pixels*, isto faz com que a detecção de movimento seja errônea. Com isto, tanto os parâmetros de deslocamento quanto os parâmetros de escala não são verdadeiros. Isto é possível afirmar se compararmos o resultado com outros autores, o que será feito na seção seguinte.

Analizando ainda as seqüências *silent* e *hall*, apesar delas não terem movimentação de câmera, foram encontrados valores tanto para deslocamento quanto para escala. Para os parâmetros de deslocamento é possível afirmar que se trata da interferência da movimentação dos objetos fora do quadro central desconsiderado para os vetores de movimento. Esta pequena influência também é sentida nos parâmetros de escala.

Apesar dos erros serem eventuais, e a variação detectada ser pequena, os erros propagados por todos os quadros deixa o resultado final bem distante do esperado. Uma possível solução é não detectar parâmetros de deslocamento caso o vídeo não apresente movimentação de câmera. Esta implementação fica como proposição para trabalhos futuros.

4.6. Comparação com outros métodos

Para verificar se a metodologia proposta é eficiente, serão feitas algumas comparações com o método proposto por Sikora [9] e os resultados obtidos por ele.

Este autor utiliza a estimação global de movimento dividida em três etapas. Cada etapa tem como função refinar os parâmetros de movimento e aumentar gradativamente a complexidade do modelo de movimento global. Na primeira etapa é estimada a translação entre o quadro atual e o quadro estimado do mosaico, através de BMA. Na etapa seguinte são estimados os parâmetros de zoom, rotação e os parâmetros de translação são refinados. A etapa três é semelhante à etapa dois, porém com o modelo de transformação parabólico, que leva em conta transformações de mais alta ordem.

A etapa dois do algoritmo do Sikora é um processo que fica em *loop* até a convergência do melhor casamento da imagem transformada pelos parâmetros detectados na etapa anterior. Isto provoca um custo computacional elevado.

Diferentemente do utilizado por Sikora, este trabalho utiliza o casamento de pontos característicos e não o casamento de blocos.

A fim de mostrar a semelhança dos resultados obtidos são apresentadas a seguir as estimativas dos parâmetros obtidos por Sikora para a seqüência *stefan* e *bus*. São feitas comparações com os obtidos neste trabalho.

A Figura 39 mostra o deslocamento horizontal para a seqüência *stefan* obtido por Sikora.

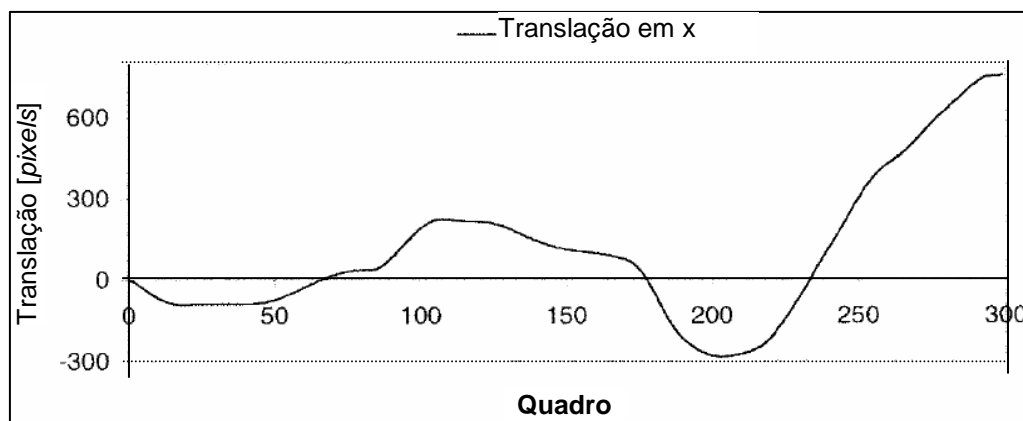


Figura 39. Translação em x da seqüência *stefan* por Sikora.

A Figura 40 apresenta a evolução do parâmetro de translação x da seqüência *stefan* obtidos neste trabalho.

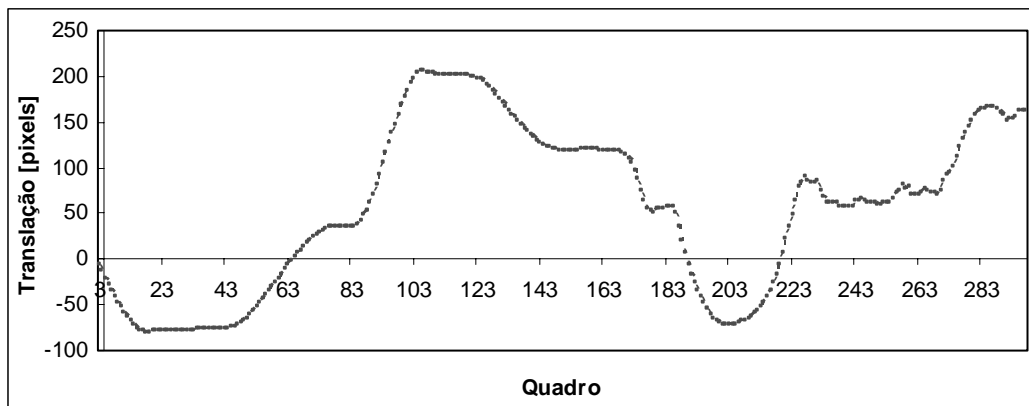


Figura 40. Translação em x da sequência *stefan* detectado.

Podemos notar a grande semelhança nos resultados obtidos nos dois trabalhos para a sequência *stefan*. Apenas entre os quadros 225 e 265 que ocorre alguma diferença, mas é devido ao deslocamento horizontal da sequência ser maior que a janela máxima considerada, como já foi descrito anteriormente.

A Figura 41 apresenta o resultado obtido para a translação horizontal da sequência *bus* obtida por Sikora.

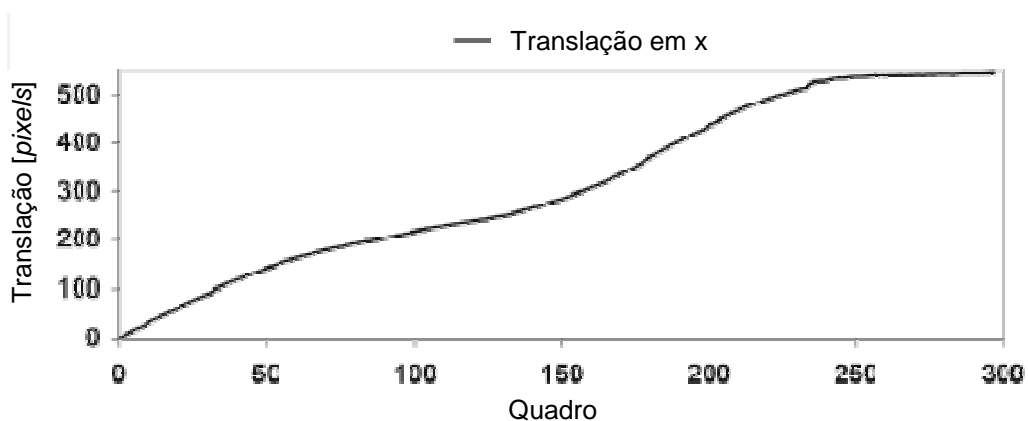


Figura 41. Translação em x da sequência *bus* por Sikora.

A Figura 42 mostra a evolução do parâmetro de translação horizontal da sequência *bus* obtido neste trabalho.

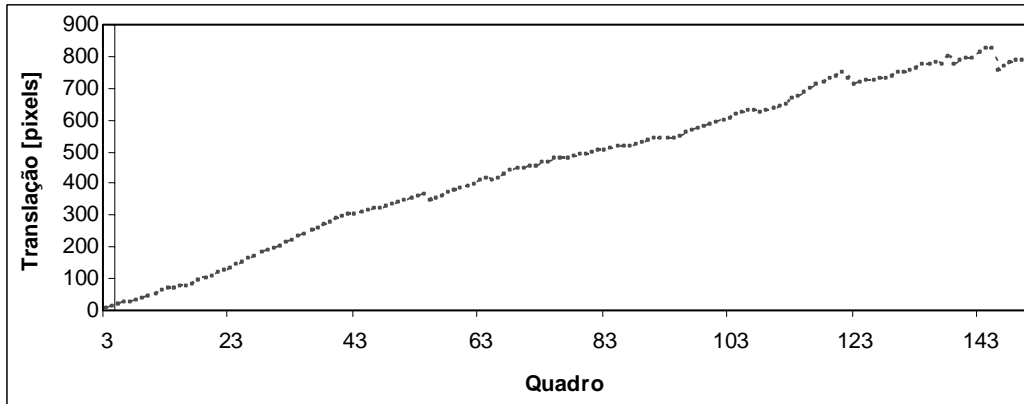


Figura 42. Translação em x da sequência *bus* detectado.

A Figura 43 apresenta os resultados para os parâmetros de escala horizontal e vertical obtidos por Sikora.

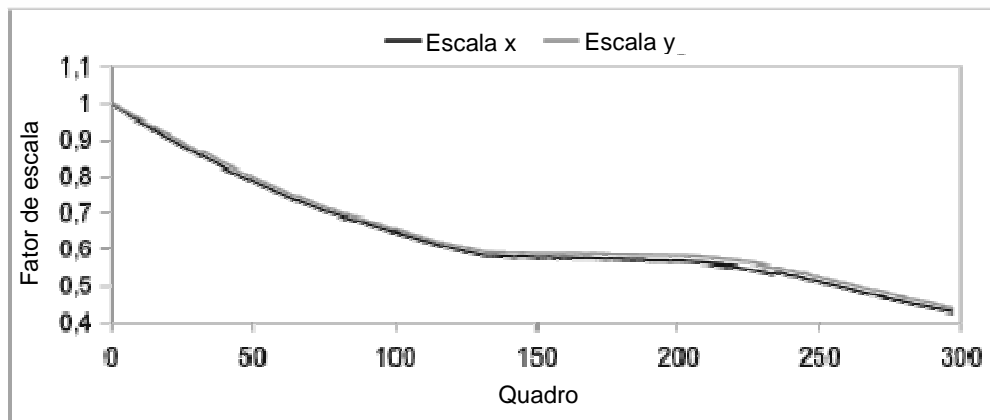


Figura 43. Escala em x e y da sequência *bus* por Sikora.

A Figura 44 mostra os mesmos parâmetros para esta metodologia apresentada.

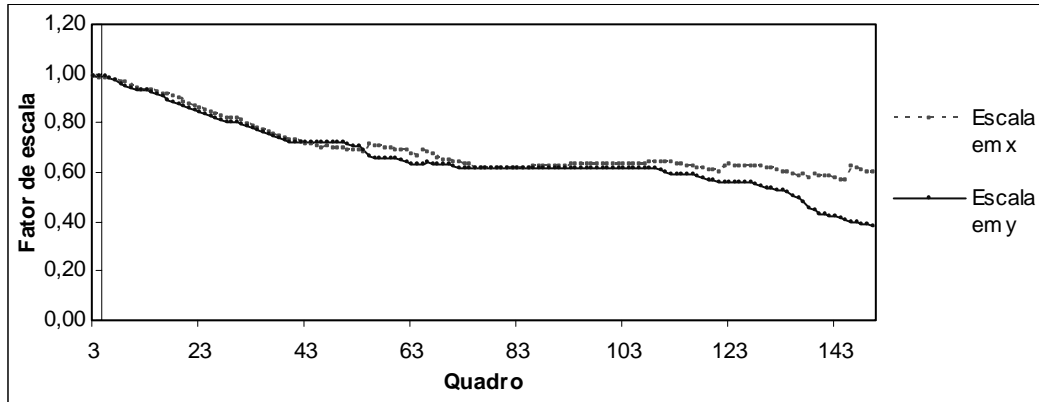


Figura 44. Escala em x e y da sequência *bus* detectados.

Deve ser observado que o vídeo utilizado por Sikora continha uma taxa de atualização de 60 quadros por segundo e que a sequência analisada neste trabalho tem uma taxa de atualização de 30 quadros por segundo. Isto provoca uma maior precisão na obtenção dos resultados por Sikora.

A semelhança dos resultados nas duas metodologias é notável, mostrando que a metodologia proposta atende as expectativas de determinação dos parâmetros de movimento da câmera.

4.7. Custo Computacional

Será mostrada nesta seção uma comparação do custo computacional envolvido na obtenção dos parâmetros de movimento global da metodologia proposta neste trabalho com as metodologias que utilizam do BMA e também do gradiente descendente.

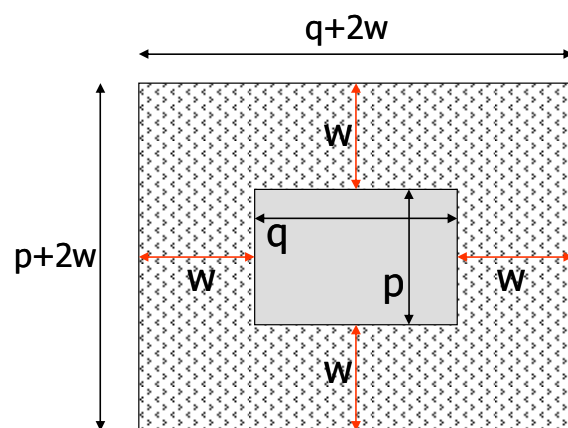


Figura 45. Janela de comparação do BMA

O algoritmo BMA busca o melhor casamento entre a janela de comparação ($p \times q$) com cada posição candidata da janela de busca $((q+2w) \times (p+2w))$ mostrada na Figura 45, isto é, existe um total de $(2w+1)^2$ posições possíveis a serem examinadas. A maior similaridade da o melhor casamento e consequentemente a posição do deslocamento do bloco.

Para uma janela de comparação de 16×16 pixels ($p = q = 16$) e uma janela de busca de 16×16 pixels ($w = 16$) é possível calcular o número de comparações que serão necessários para o método BMA, mostrado na Tabela 1.

Tabela 1. Número de comparações do BMA

Método BMA	Nº de comparações
Busca completa	256
Busca em 3 passos	25
Busca em diamante	13
Busca logarítima 2D	15

Isto indica que se na metodologia proposta, o número de pontos comparados forem de até 12 pontos, ainda teríamos um ganho computacional na etapa de casamento de blocos, pois seria mais rápido que a busca em diamante que tem o menor número de comparações para os tipos de busca do algoritmo BMA. Deve-se levar em consideração que estas quantidades de comparações mostradas na Tabela 1 são referentes ao pior caso de cada tipo de busca.

Foi verificado que, para a grande maioria dos pontos, são feitas em média 5 comparações com outros pontos, ou seja, em uma janela de busca, apenas 6 pontos em média são encontrados. Isto mostra que o ganho computacional é considerável.

Foram feitos testes em um computador “Celeron 2,0 GHz com 512 MB de memória” e verificou-se que, para as seqüências testadas, era necessário 1,1 segundos para se fazer a análise em cada quadro utilizando o BMA. Substituindo na metodologia a busca BMA pelo casamento de pontos característicos, houve um ganho em velocidade de 20%, 19,4%, 15% e 10% para as seqüências *bus*, *stefan*, *hall* e *silent* respectivamente.

A metodologia proposta por Dufaux, que utiliza o gradiente descendente na estimação do movimento global, tem uma etapa de busca por BMA e uma etapa iterativa que necessita de convergência. Isto mostra que em comparação a Dufaux, a metodologia proposta neste trabalho também tem uma diminuição do custo computacional.

5. Conclusão

Neste trabalho foi apresentada a proposição de uma nova metodologia de estimação do movimento global de uma seqüência de vídeo utilizando a movimentação de pontos característicos entre quadros subseqüentes.

Analisando os resultados podemos afirmar que os mesmos ficaram próximos do esperado. É possível afirmar a partir do momento que comparamos os resultados obtidos com os resultados de outros autores. Os resultados para as seqüências sintetizadas mostram que é possível a utilização desta metodologia para determinação do movimento global, uma vez que os erros encontrados são pequenos.

Apesar dos parâmetros encontrados para as seqüências sintetizadas não serem os esperados, são parâmetros que solucionam o problema da estimação do movimento global. Isto é devido à equação que representa a transformação não ser unívoca, o que indica que existe mais de uma solução para o problema, demonstrando que a metodologia não prioriza cada movimentação em particular, mas sim o conjunto das movimentações.

É uma metodologia de rápida implementação e por não possuir nenhum processo iterativo pode ser considerada como tendo um ganho computacional em relação às outras propostas.

São encontrados pequenos erros em seqüências sem movimentação de câmera. Isto é devido a erros introduzidos por ruídos presentes na imagem e também pela influência de partes do objeto que não se encontram no centro da imagem, sendo assim considerado na análise.

Também é verificado que existe uma propagação de pequenos erros devido principalmente à imprecisão da medição de pequenos deslocamentos locais próximos a 1 *pixel*. A solução proposta é utilizar os parâmetros propagados para no máximo 12 quadros, como é utilizado nas codificações MPEG-1 e 2.

É uma metodologia que atende a praticamente todas as transformações que são encontradas nos vídeos. Deve ser levado em conta que transformações que provoquem deslocamentos maiores que a janela de busca

dos pontos considerada geram erros na análise. Caso isto ocorra, basta aumentar a janela. Porém, isto introduz um custo a mais na computação dos vetores locais, pois haverá uma busca maior entre os pontos dentro desta janela.

Tendo em vista que os principais objetivos deste trabalho foram alcançados, uma metodologia de baixo custo computacional e com eficiente detecção do movimento global, fica a proposição para sua utilização nos padrões de codificação que necessitam da determinação do movimento global ou a movimentação da câmera de aquisição.

5.1. Trabalhos futuros

Algumas sugestões para trabalhos e estudos futuros são:

- Desenvolvimento de uma etapa interativa para melhorar a detecção do movimento global, porém sem muito aumento do custo computacional;
- Melhoria da validação dos vetores de movimento local utilizando a exclusão do objeto e não de uma região central;
- Aproveitamento da metodologia para a segmentação de objetos;
- Aproveitamento da metodologia para obtenção da imagem panorâmica do vídeo (*sprite, mosaico*);
- Avaliação do desempenho da metodologia através da utilização em um codificador MPEG-4.
- Alteração do algoritmo para aumento da velocidade de processamento para utilização em tempo real.

6.Referências Bibliográficas

- [1] L. Chiariglione, "Short MPEG-1 description",
<http://www.chiariglione.org/mpeg/standards/mpeg-1/mpeg-1.htm>, Janeiro 2005.
- [2] L. Chiariglione, "Short MPEG-2 description",
<http://www.chiariglione.org/mpeg/standards/mpeg-2/mpeg-2.htm>, Janeiro 2005.
- [3] R. Koenen, "MPEG-4 Overview – (V.21 – Jeju Version)",
<http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm>, Janeiro 2005.
- [4] F. Dufaux, J. Konrad, *Efficient, Robust and Fast Global Motion Estimation for Video Coding*. IEEE Transactions on Image Processing, vol. 9, no. 3, Mar-2000.
- [5] G. Giunta e U. Mascia, *Estimation of Global Motion Parameters by Complex Linear Regression*. IEEE Trans. Image Processing, vol. 8, pág. 1652-1657, Novembro 1999.
- [6] M. Kim, J. G. Choi, D. Kim, H. Lee, M. H. Lee, C. Ahn, Y.-S. Ho. *A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences Based on Spatio-Temporal Information*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 8. Dec-99.
- [7] T. Chen, W. Lin e C. Chen, *Artificial Neural Networks for 3-D Motion Analysis –Part I: Rigid Motion*. IEEE Trans. Neural Networks, vol. 6, pág. 1386-1393, Novembro 1995.
- [8] S. Goldenstein, *Statistical Cue Estimation for Model-Based Shape and Motion Tracking*, Dissertation in Computer and Information Science, University of Pennsylvania, 2002.
- [9] A. Smolic, T. Sikora, and J.-R. Ohm, *Long-Term Global Motion Estimation and Its Application for Sprite Coding, Content Description, and Segmentation*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 8. Dec-1999.
- [10] C. R. Wren, *Perspective Transform Estimation*.
<http://xenia.media.mit.edu/~cwren/interpolator/>, Janeiro 2005.
- [11] F. Zhu, P. Xue e E. Ong, *Low-Complexity Global Motion Estimation Based on Content Analysis*. ISCAS '03, vol. 2 , pág. 624-627, Maio 2003.

- [12] E. T. Kim e H. Kim, *Fast and Robust Parameter Estimation Method for Global Motion Estimation in the Video Coder*. IEEE Trans. Consumer Electronics, vol. 45, pág. 76-83, Fevereiro 1999.
- [13] D. Wang e L. Wang, *Global Motion Parameters Estimation Using a Fast and Robust Algorithm*. IEEE Trans. Circuits and Syst. Video Technol., vol. 7, pág. 823-826, Outubro 1997.
- [14] H. Cheung e W. Siu, *Fast Global Motion Estimation for Sprite Generation*. ISCAS 2002, vol. 3, pág. 1-4, Maio 2003.
- [15] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge Univ. Press, 1988.
- [16] S. Ghael, G. Chew, *Creating Image Mosaics*. EE392J – Digital Video Processing, Winter '00.
- [17] M.V.Lamar, J.C.Bermudez e R.Seara, *Interframe Image Coding Using QuadTree for Motion Estimation*, Anais do 38th Midwest Symposium on Circuits and Systems, pp. 862-865, Rio de Janeiro, RJ, Agosto 1995.
- [18] C. Harris and M. Stephens, *A combined corner and edge detector*, Fourth Alvey Vision Conference, pp.147-151, 1988.
- [19] A. Criminisi, I. Reid, A. Zisserman, *A Plane Measuring Device*. Department of Engineering Science, University of Oxford, Oxford, UK.
www.robots.ox.ac.uk/~vgg/presentations/bmvc97/criminispaper/planedev.html
- [20] M. Sonka, V. Hlavac, R. Boyle. *Image Processing, Analysis, and Machine Vision*. 2nd Edition. PWS Publishing, 1999. P. 186-188.
- [21] A. Neri, S. Colonnese, G. Russo. *Video sequence segmentation for object-based coders using higher order statistics*. IEEE International Symposium on Circuit and Systems. Jun-97.
- [22] C. Kim, J.-N. Hwang. *Fast and Automatic Video Object Segmentation and Tracking for Content-Based Applications*. IEEE Transactions on Circuits and Systems for Video Technology, vol.12, no. 2. Feb-2002.
- [23] M. Pollefeys, *Tutorial on 3D Modeling from Images*, Lecture notes, Katholieke Universiteit Leuven, Jun-2000.